

# Leveraging the L1: The role of EAL learners' first language in their acquisition of English vocabulary

Hamish William Chalmers

Thesis submitted in partial fulfilment of the requirements of the award of Doctor of Philosophy

Oxford Brookes University

February 2019

# Declaration

I hereby certify that this dissertation constitutes my own product, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions or writings of another.

I declare that the dissertation describes original work that has not previously been presented for the award of any other degree of any institution.

Hamish W Chalmers

# ABSTRACT

Approximately one in five primary school pupils in England are classified as having English as an additional language (EAL), meaning that they routinely use, or are exposed to, languages other than English. It is commonly thought that EAL learners' first languages (L1s) can be leveraged to positively impact their linguistic and academic development in English. However, despite an abundance of theoretical and observational evidence used to argue this position, there is little experimental evidence to clarify the extent and nature of any relevant causal relationships. This gap in evidence was revealed in the first original contribution of this thesis: a systematic review of experimental research on the educational effects of mediating primary and pre-primary multilingual learners' learning through their L1.

An extensive search of twelve bibliographic databases revealed only ten studies that met the review's inclusion criteria. The pedagogical focuses of these studies varied, and the overall picture was unclear. However, five studies coalesced around the use of L1 as a mediating tool for teaching English vocabulary. Of these, three found that L1-mediation was associated with improved knowledge of the target vocabulary, one found an advantage associated with not using the L1, and the remaining study had mixed results. None of these studies were conducted with linguistically diverse groups of students, typical in English schools.

The promising, if somewhat tentative, conclusion invited by the systematic review informed the second original contribution of this thesis: a randomised crossover trial with linguistically diverse students, comparing the effects of L1-mediated teaching and English-only teaching on English vocabulary learning. Forty EAL learners aged from 8 to 11 years, representing 14 different L1s took part. Participants watched short videos that explained the meanings of items of vocabulary taken from the National Curriculum for England. The spoken content of the videos was either in each participant's L1 or in English. Analysis of the expressive and receptive knowledge of the target words following each condition revealed no statistically significant differences in outcomes.

The implications for pedagogy and future research are discussed, especially the imperative to conduct more and better research to add clarity to our understanding of the causal relationships between different types of L1-mediation and linguistic and academic success in linguistically diverse classrooms.

## Acknowledgements

I could not have done this without the unwavering love and support - intellectual, moral, material – of Iain and Jan Chalmers. Words can't express what you mean to me. To Map Chalmers, for following me 5,952 miles around the world so that I could pursue this folly – I love you. Millie and Layla Chalmers, you've kept me focused on the important things – so much love! Theo Chalmers, brother in arms.

This thesis would not have been possible without Vicki Murphy. You took me in and kept me pointed in the right direction. You didn't have to. Your generosity of spirit is unmatched. I want to be you when I grow up.

My journey to this point has been guided from the start by powerful mentors. Before Vicki, the formative impact of Susan Warby and Hazel Gibbons, who both taught me how to be a teacher, still ripples in everything I do.

Thank you to the supervisors who have come and gone during the four years that it took to complete this project. Adroula Yiakoumetti, James Percival, Graham Butt, Ana Souza, Nicoleta Gaciu, and Roger Dalrymple. Quite a team! Thanks also to Annie Haight and Jane Spiro at Brookes for your consistent moral support.

Thank you to the army of translators and voice talents who gave freely of their time and expertise to make it possible to produce the 270 individual L1 resources used in this project.

And thank you to the four schools, their headteachers, teachers, support staff and students for allowing me to come in to poke around under the hood of your teaching and learning. I promised you I wouldn't name you, so I won't. But it goes without saying that education research relies on the generosity of people like you, who beckon us down from our ivory towers and into your schools so that we can learn from you as we try to understand how to make teaching better. Keep this attitude alive, but always ask to see the trial protocol.



# TABLE OF CONTENTS

<b>DECLARATION</b> .....	<b>1</b>
<b>ABSTRACT</b> .....	<b>2</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>3</b>
<b>LIST OF FIGURES</b> .....	<b>10</b>
<b>LIST OF TABLES</b> .....	<b>11</b>
<b>CHAPTER 1 INTRODUCTION</b> .....	<b>13</b>
<b>1.1 STATEMENT OF THE PROBLEM</b> .....	<b>13</b>
<b>1.2 WHY IS THE ISSUE OF L1 USE AMONG CHILDREN WHO USE ENGLISH AS AN ADDITIONAL LANGUAGE AN ISSUE THAT DESERVES ATTENTION?</b> .....	<b>16</b>
<b>1.3 WHO ARE EAL PUPILS?</b> .....	<b>16</b>
<b>1.4 AIM OF THE THESIS</b> .....	<b>18</b>
<b>CHAPTER 2 LITERATURE REVIEW</b> .....	<b>20</b>
<b>2.1 INTRODUCTION</b> .....	<b>20</b>
<b>2.2 OVERVIEW OF RELEVANT RESEARCH</b> .....	<b>20</b>
<b>2.3 A BIT OF HISTORY</b> .....	<b>21</b>
2.3.1 FROM GRAMMAR TRANSLATION TO SECOND LANGUAGE SUBMERSION.....	23
2.3.2 THE REHABILITATION OF L1 AS A NECESSARY COMPONENT IN EFFECTIVE LANGUAGE LEARNING .....	25
<b>2.4. OVERVIEW OF THE THEORETICAL POSITION</b> .....	<b>26</b>
2.4.1 COMMON UNDERLYING PROFICIENCY .....	26
2.4.2 BASIC INTERPERSONAL COMMUNICATIVE SKILLS AND COGNITIVE ACADEMIC LANGUAGE PROFICIENCY .....	27
2.4.3 THRESHOLD HYPOTHESIS .....	28
<b>2.5 PEDAGOGICAL IMPLICATIONS OF THE THEORY</b> .....	<b>29</b>
2.5.1 USING THE L1 TO ACTIVATE PRIOR SUBJECT KNOWLEDGE .....	29
2.5.2 USING THE L1 AS A BASIS FOR LINGUISTIC DEVELOPMENT .....	30
<b>2.6 WHAT KINDS OF RESEARCH HELPS US TO UNDERSTAND THE POTENTIAL FOR USING THE L1 IN THE L2 CLASSROOM?</b> .....	<b>31</b>
2.6.1 RESEARCH ON CROSS LINGUISTIC TRANSFER .....	32
2.6.2 TRANSFER OF LOWER ORDER SKILLS .....	32
2.6.3 TRANSFER OF HIGHER ORDER SKILLS .....	34
2.6.4 TRANSFER ACROSS DISSIMILAR LANGUAGES.....	35
2.6.5 LONG-TERM EFFECTS ATTRIBUTED TO CROSS-LINGUISTIC TRANSFER .....	36
2.6.6 SUMMARY OF RESEARCH ON CROSS-LINGUISTIC TRANSFER .....	36
<b>2.7 EVIDENCE FROM BILINGUAL PROGRAMMES OF EDUCATION</b> .....	<b>37</b>
2.7.1 THE ST. LAMBERT EXPERIMENT: EARLY EVIDENCE FOR THE EFFECTS OF BILINGUAL SCHOOLING .....	37
2.7.2 SYSTEMATIC REVIEWS OF RESEARCH IN BILINGUAL PROGRAMMES OF EDUCATION .....	39
2.7.3 REVIEW OF BILINGUAL READING PROGRAMMES BY SLAVIN AND CHEUNG (2005).....	41
2.7.4 REVIEW OF BILINGUAL PROGRAMMES BY ROLSTAD, MAHONEY AND GLASS (2005).....	43
2.7.5 REVIEW OF BILINGUAL PROGRAMMES IN EUROPE BY RELJIĆ, FERRING AND MARTIN (2015) .....	44

2.7.6 RANDOMISED TRIAL OF DISTRICT-WIDE BILINGUAL PROGRAMMES.....	45
2.7.7 SUMMARY OF RESEARCH ON BILINGUAL EDUCATION .....	47
<b>2.8 HOW DO MULTILINGUAL LEARNERS USE THEIR L1S AT SCHOOL?.....</b>	<b>48</b>
2.8.1 USING THE L1 FACILITATES CO-OPERATION AND LIGHTENS THE COGNITIVE LOAD .....	48
2.8.2 ... BUT DOES MORE FLUID CO-OPERATION AND A LIGHTENED COGNITIVE LOAD LEAD TO BETTER TASK PERFORMANCE? .	49
2.8.3 SUMMARY OF RESEARCH ON HOW MULTILINGUAL LEARNERS USE THEIR L1 IN SCHOOL.....	51
<b>2.9 TRANSLATING RESEARCH INTO THE UK CONTEXT .....</b>	<b>51</b>
2.9.1 LIMITATIONS OF THE RESEARCH .....	52
2.9.2 UNPACKING THE BLACK BOX OF BILINGUAL EDUCATION .....	53
<b>2.10 STAND-ALONE STRATEGIES THAT USE THE L1 AS A PEDAGOGICAL TOOL .....</b>	<b>57</b>
2.10.1 OBSERVATIONAL RESEARCH .....	57
2.10.2 USING L1 TO PREPARE FOR L2 ACTIVITIES .....	58
2.10.3 SYSTEMATIC COMPARISON OF FEATURES OF L1 WITH L2 .....	59
2.10.4 USING THE L1 TO GIVE INSTRUCTIONS AND TO TEACH ACADEMIC SKILLS.....	60
2.10.5 USING THE L1 TO TEACH L2 VOCABULARY .....	62
2.10.6 SUMMARY OF RESEARCH ON STAND-ALONE STRATEGIES FOR USING L1.....	64
<b>2.11 SUMMARY AND CONCLUSION .....</b>	<b>64</b>
<b>CHAPTER 3 SYSTEMATIC REVIEW .....</b>	<b>69</b>
<b>3.1 INTRODUCTION.....</b>	<b>69</b>
3.1.1 WHY A 'SYSTEMATIC' REVIEW? .....	69
3.1.2 THE FOCUS OF THIS SYSTEMATIC REVIEW .....	71
<b>3.2 STRUCTURED SUMMARY.....</b>	<b>71</b>
3.2.1 BACKGROUND .....	71
3.2.2 OBJECTIVES.....	71
3.2.3 METHODS.....	72
3.2.4 RESULTS .....	72
3.2.5 DISCUSSION .....	72
<b>3.3 AIMS OF THE SYSTEMATIC REVIEW.....</b>	<b>73</b>
3.3.1 REVIEW QUESTIONS.....	73
<b>3.4 METHODS .....</b>	<b>74</b>
3.4.1 PROTOCOL REGISTRATION .....	74
3.4.2 INFORMATION SOURCES.....	74
3.4.3 SEARCH .....	75
3.4.4 ELIGIBILITY CRITERIA.....	78
3.4.5 LANGUAGE OF PUBLICATION .....	78
3.4.6 STUDY SELECTION .....	82
3.4.6.1 Phase 1: screening titles and abstracts.....	82
3.4.6.2 Phase 2: locating full reports .....	82
3.4.6.3 Phase 3: citation searching .....	83
3.4.6.4 Quality assurance .....	83
3.4.7 DATA EXTRACTION .....	83
3.4.8 RISK OF BIAS ASSESSMENT .....	84
3.4.9 CHOOSING AN APPROPRIATE RISK OF BIAS ASSESSMENT TOOL .....	85
3.4.10 ADJUSTING THE EPHPP TOOL FOR USE WITH EDUCATION RESEARCH.....	86
3.4.10.1 Terminology.....	87

3.4.10.2 <i>Sample representativeness</i> .....	87
3.4.10.3 <i>Control of confounders</i> .....	89
3.4.11 INCOMPLETE REPORTING.....	90
3.4.12 SUMMARY MEASURE.....	90
3.4.13 SYNTHESIS OF RESULTS .....	90
3.4.14 RISK OF BIAS ACROSS STUDIES .....	91
<b>3.5 RESULTS .....</b>	<b>91</b>
3.5.1 STUDY SELECTION .....	91
3.5.2 INCLUDED STUDIES.....	94
3.5.3 STUDY CHARACTERISTICS.....	95
3.5.3.1 <i>Location</i> .....	95
3.5.3.2 <i>Publication status</i> .....	95
3.5.3.3 <i>Design</i> .....	95
3.5.3.4 <i>Size</i> .....	95
3.5.3.5 <i>Duration</i> .....	95
3.5.3.6 <i>Educational programme type</i> .....	96
3.5.3.7 <i>Language/Skills focus</i> .....	96
3.5.3.8 <i>Findings</i> .....	96
3.5.4 RISK OF BIAS ASSESSMENT WITHIN STUDIES.....	106
3.5.4.1 <i>Selection bias</i> .....	106
3.5.4.2 <i>Study design</i> .....	106
3.5.4.3 <i>Confounders</i> .....	107
3.5.4.4 <i>Allocation concealment, blinding of outcome assessment and blinding of participants to research questions</i> .....	107
3.5.4.5 <i>Data collection</i> .....	108
3.5.4.6 <i>Withdrawals and dropout</i> .....	108
3.5.5 QUALITY ASSESSMENT ACROSS STUDIES.....	110
3.5.6 INDIVIDUAL STUDY SUMMARIES .....	110
1. <i>Effects of parent volunteers reading first language (L1) books to ESL students</i> .....	111
2. <i>Effects of Parent Participation Using First Language Curriculum Materials on the English Reading Achievement and Second-Language Acquisition of Hispanic Students</i> .....	113
3. <i>A Bidialectal Programme for the Learning of Standard Modern Greek in Cyprus</i> .....	116
4. <i>A possible role for the first language in young learners' processing and storage of foreign language vocabulary</i> .....	119
5. <i>L2 Vocabulary Acquisition in Children: Effects of Learning Method and Cognate Status</i> .....	122
6. <i>Facilitating vocabulary acquisition of young English language learners</i> .....	124
7. <i>The Cross-Linguistic Effects of Dialogic Reading on Young English Language Learners</i> .....	126
8. <i>Investigating age in the use of L1 or English-only instruction: Vocabulary acquisition by Korean EFL learners</i> .....	129
9. <i>Harnessing linguistic diversity in polylingual British curriculum schools. Do L1 mediated home learning tasks improve learning outcomes for bilingual children? A randomised trial</i> .....	131
10. <i>The Effects of Using L1 Translation on Young Learners' Foreign Language Vocabulary Learning</i> ..	133
3.5.7 SYNTHESIS OF RESULTS AND BOTTOM-LINE FINDINGS .....	136
<b>3.6 DISCUSSION .....</b>	<b>137</b>
3.6.1 PEDAGOGIC FOCUS .....	137
3.6.1.1 <i>Vocabulary</i> .....	137
3.6.1.2 <i>General reading proficiency</i> .....	139
3.6.1.3 <i>Oral and written proficiency in the L2</i> .....	140
3.6.2 LOCATION AND TARGET POPULATION.....	141
3.6.3 CONCLUSIONS BASED ON THE QUALITY OF EVIDENCE (BEST EVIDENCE SYNTHESIS) .....	142
3.6.3.1 <i>Randomised trials</i> .....	142

3.6.3.2. <i>Non-randomised comparisons</i> .....	142
3.6.3.4 <i>Designs with no comparison group</i> .....	143
3.6.3.5 <i>Summary</i> .....	143
<b>3.7 CONCLUSIONS</b> .....	<b>144</b>
3.7.1 THE NATURE OF THE RESEARCH .....	144
3.7.1.1 <i>Skills focus</i> .....	144
3.7.1.2 <i>Location and languages</i> .....	145
3.7.1.3 <i>Methods</i> .....	145
3.7.2 EXTENT OF THE RESEARCH .....	145
3.7.3 IMPLICATIONS FOR PRACTICE.....	145
3.7.4 IMPLICATIONS FOR FUTURE RESEARCH .....	146
3.7.6 LIMITATIONS OF THIS SYSTEMATIC REVIEW .....	147
3.7.6.1 <i>Electronic searching, hand searching, and language of publication</i> .....	147
3.7.6.2 <i>Grey literature and publication bias</i> .....	147
3.7.6.3 <i>Quality assurance</i> .....	147
<b>3.8. INFORMING THE DIRECTION OF THE NEXT STAGE OF THIS DOCTORAL PROJECT</b> .....	<b>148</b>
3.8.1 DECIDING THE FOCUS OF THE INTERVENTION STUDY.....	148
3.8.2 WARRANT FOR THE INTERVENTION STUDY .....	149
3.8.3 ACCOMMODATING LINGUISTIC DIVERSITY .....	150
3.8.4 OVERVIEW OF THE INTERVENTION STUDY .....	151
<b>CHAPTER 4 METHODOLOGY</b> .....	<b>153</b>
<b>4.1 AIMS AND OBJECTIVES OF THE RESEARCH</b> .....	<b>153</b>
4.1.2 RESEARCH QUESTIONS .....	154
4.1.3 HYPOTHESES .....	154
<b>4.2 RESEARCH PARADIGMS, ONTOLOGY AND EPISTEMOLOGY</b> .....	<b>155</b>
<b>4.3 EXPERIMENTAL DESIGN</b> .....	<b>158</b>
4.3.1 RANDOMISED CROSSOVER TRIAL.....	158
4.3.2 RATIONALE FOR THE DESIGN .....	158
4.3.3 LIMITATIONS OF CROSSOVER DESIGNS.....	159
<b>4.4. SETTING OF THE STUDY, SCHOOLS AND PARTICIPANTS</b> .....	<b>160</b>
4.4.1 THE CITY OF OXFORD .....	160
4.4.2 THE SCHOOLS .....	161
4.4.2.1 <i>Sampling frame and sampling technique</i> .....	161
4.4.2.2 <i>Finding eligible schools</i> .....	161
4.4.2.3 <i>Sample size</i> .....	162
4.4.2.4 <i>Recruiting schools</i> .....	163
4.4.2.5 <i>Characteristics of participating schools</i> .....	164
4.4.3 PARTICIPANTS.....	166
4.4.3.1 <i>Recruiting participants</i> .....	166
4.4.3.2 <i>Participants</i> .....	169
4.4.3.3 <i>Language proficiency of the participants</i> .....	171
4.4.3.4 <i>Procedure for carrying out the language background questionnaire</i> .....	172
4.4.3.5 <i>Summary of the information collected using the language background questionnaire</i> .....	173
4.4.3.6 <i>Linguistic and educational background</i> .....	174
4.4.3.7 <i>English language proficiency</i> .....	176
<b>4.5 ALLOCATION TO COMPARISON CONDITIONS</b> .....	<b>177</b>

<b>4.6 INTERVENTIONS .....</b>	<b>178</b>
4.6.1 OVERVIEW .....	178
4.6.2 PROCEDURE .....	179
4.6.3 MATERIALS .....	180
4.6.3.1 <i>The Frayer model of concept mastery</i> .....	180
4.6.3.2 <i>Introducing the concept maps to the participants</i> .....	182
4.6.3.3 <i>Selecting the target words</i> .....	182
4.6.3.4 <i>Creating and standardising the video voiceover scripts</i> .....	185
4.6.3.5 <i>Translating and recording the voiceover scripts</i> .....	187
4.6.3.7 <i>Creating the videos</i> .....	188
4.6.3.8 <i>Making the videos available for the intervention</i> .....	188
4.6.4 INSTRUCTIONAL ORDER.....	190
4.6.5 ASSESSMENTS.....	192
<b>4.7 TESTING THE MATERIALS AND PROCEDURE.....</b>	<b>194</b>
<b>4.8 METHODS TO ANALYSE DATA.....</b>	<b>195</b>
<b>4.9 ETHICS.....</b>	<b>195</b>
4.9.1 THE ETHICAL WARRANT FOR THE STUDY .....	195
4.9.2 INFORMED CONSENT .....	195
4.9.3 RISK MANAGEMENT AND DATA PROTECTION .....	196
4.9.4 REPORTING FINDINGS TO PARTICIPANTS .....	197
<b>CHAPTER 5 RESULTS.....</b>	<b>198</b>
<b>5.1 INTRODUCTION.....</b>	<b>198</b>
<b>5.2 RESTATEMENT OF THE RESEARCH QUESTIONS AND METHODS OF ANALYSIS.....</b>	<b>198</b>
<b>5.3 DATA ANALYSIS.....</b>	<b>199</b>
5.3.1 RESEARCH QUESTION 1 .....	199
5.3.1.1 <i>Descriptive statistics of the results of the post-test</i> .....	199
5.3.1.2 <i>Distribution</i> .....	200
5.3.1.3 <i>Inferential Statistics</i> .....	201
5.3.2 CHANGE IN TARGET VOCABULARY KNOWLEDGE.....	202
5.3.2.1 <i>Descriptive statistics</i> .....	202
5.3.2.2 <i>Inferential statistics</i> .....	203
<b>5.4 SUMMARY OF FINDINGS FOR RQ1.....</b>	<b>204</b>
<b>5.5 RESEARCH QUESTION 2 .....</b>	<b>204</b>
5.5.1 CORRELATIONS .....	205
<b>5.6 SUMMARY OF FINDINGS OF RQ2.....</b>	<b>206</b>
<b>5.7 SUMMARY OF ALL RESULTS.....</b>	<b>206</b>
<b>CHAPTER 6 DISCUSSION .....</b>	<b>207</b>
<b>6.1 INTRODUCTION.....</b>	<b>207</b>
<b>6.2 CONTEXT AND AIM: REVISITED.....</b>	<b>207</b>
6.2.1 USE OF THE L1: THEORY AND OBSERVATIONAL EVIDENCE.....	208
6.2.2 USE OF THE L1: NEUROCOGNITIVE EXPLANATIONS FOR WHY THIS MIGHT BE HELPFUL TO MULTILINGUAL LEARNERS ...	209
6.2.3 USE OF THE L1: EVIDENCE FROM EXPERIMENTS .....	209
6.2.4 SUMMARY OF THE LITERATURE REVIEW AND NEXT STEPS.....	210

<b>6.3 SYSTEMATIC REVIEW OF INTERVENTION STUDIES .....</b>	<b>210</b>
6.3.1 SUMMARY OF FINDINGS .....	210
6.3.2 IMPLICATIONS FROM THE SYSTEMATIC REVIEW .....	211
<b>6.4 INTERVENTION STUDY.....</b>	<b>211</b>
<b>6.5 FINDINGS OF THE INTERVENTION STUDY.....</b>	<b>212</b>
6.5.1 FINDINGS FOR RQ1: .....	212
6.5.2 FINDINGS FOR RQ2: .....	213
6.5.2.1 <i>Associations between high English proficiency and recall of words taught in both L1 and English-only conditions</i> .....	214
6.5.2.2 <i>In the L1 condition better L1 proficiency is associated with better word learning</i> .....	215
6.5.3 A NOTE ON VARIANCE.....	215
<b>6.6 IMPLICATIONS FOR PRACTICE, RESEARCH AND THEORY .....</b>	<b>216</b>
6.6.1 PRACTICE – USE OF THE L1 IS A PERSONAL CHOICE .....	216
6.6.2 RESEARCH – MORE AND BETTER RESEARCH IS NEEDED .....	220
6.6.2.1 <i>Volume of relevant research</i> .....	220
6.6.2.2 <i>Quality of relevant research</i> .....	221
6.6.2.3 <i>Clear reporting of research</i> .....	221
6.6.2.4 <i>Open science practices</i> .....	223
6.6.3 THEORY - A PROBLEM WITH THE ORTHODOXY?.....	223
<b>6.7 LIMITATIONS OF THE STUDY.....</b>	<b>225</b>
6.7.1 RECRUITING A REPRESENTATIVE SAMPLE OF THE POPULATION OF EAL LEARNERS .....	225
6.7.1.1 <i>'Hard to Reach' Parents</i> .....	226
6.7.1.2 <i>EAL, stigma and immigration</i> .....	227
6.7.2 SELECTION BIAS.....	228
6.7.3 ACCURATE LANGUAGE PROFICIENCY ASSESSMENT .....	230
6.7.3.1 <i>Proficiency in English</i> .....	230
6.7.3.2 <i>Proficiency in L1</i> .....	232
6.7.4 SUMMARY .....	232
<b>CHAPTER 7 CONCLUSION .....</b>	<b>233</b>
<b>REFERENCES.....</b>	<b>237</b>
<b>APPENDICES.....</b>	<b>271</b>

## List of figures

Figure 2.1 "...children by the help of their mother tongue may the better learn to speak Latin in ordinary discourse." p.22

Figure 2.2 Item 15 of the founding statutes of Adam's Free Grammar School, Shropshire. p.22

Figure 3.1 PRISMA Flow diagram. p.93

Figure 3.3 Included studies by direction of outcome, size, topic and risk of bias assessment. p.136

Figure 4.1 CONSORT Flow diagram. p.168

Figure 4.2 Number of participants by school. p.169

Figure 4.3 Number of participants by year group. p.170

Figure 4.4 Number of participants by age in years. p.170

Figure 4.5 Number of participants by L1. p.170

Figure 4.6 Participants' completed years in the English education system. p.175

Figure 4.7 Participants' English language proficiency levels. p.176

Figure 4.8 Frayer model concept map. p.181

Figure 4.9 Website home page. p.189

Figure 4.10 Example of a language page on the website. p.189

Figure 4.11 Example of a set of videos on the website. p.190

Figure 4.12 Example of the expressive vocabulary test. p.193

Figure 4.13 Example of the receptive vocabulary test. p.194

Figure 6.1 The present study in context with the findings of the systematic review. p.218

## List of tables

Table 3.1 Search terms. p.77

Table 3.2 Criteria for inclusion/exclusion in the systematic review. p.79

Table 3.3 Summary of key descriptive data extracted from included studies. p.97

Table 3.4 Summary of the statistical data extracted from the included studies. p.100

Table 3.5 Risk of bias assessments across studies. p.109

Table 3.6 Risk of bias assessment for Study 1. p.112

Table 3.7 Risk of bias assessment for Study 2. p.115

Table 3.8 Risk of bias assessment for Study 3. p.118

Table 3.9 Risk of bias assessment for Study 4. p.121

Table 3.10 Risk of bias assessment for Study 5. p.123

Table 3.11 Risk of bias assessment for Study 6. p.125

Table 3.12 Risk of bias assessment for Study 7. p.127

Table 3.13 Risk of bias assessment for Study 8. p.130

Table 3.14 Risk of bias assessment for Study 9. p.132

Table 3.15 Risk of bias assessment for Study 10. p.134

Table 4.1 Self-reported level of L1 proficiency in the four language domains of Speaking, Listening, Writing and Reading. p.174

Table 4.2 School year on arrival in England. p.175

Table 4.3 Participant characteristics after random allocation to intervention order. p.178

Table 5.1 Mean scores in each intervention condition. p.200

Table 5.2 Distribution of Outcome Measures. p.200

Table 5.3 Results of paired samples t-test for expressive knowledge of the target words by condition. p.201

Table 5.5 Scores on pre- and post-tests of expressive and receptive knowledge of the target words, by condition. p.202



Table 5.6 Differences between scores on pre- and post-test expressive and receptive knowledge of the target words, by condition. p.203

Table 5.7 Difference in difference between pre- and post-test knowledge of the target words, by condition. p.204

Table 5.8 Spearman correlations between putative predictors and outcomes. p.205

Table 5.9 Results of paired samples t-test for receptive knowledge of the target words by condition for only participants with an above theoretical mean L1 proficiency rating. p.206

# CHAPTER 1 INTRODUCTION

*When the research began, primary school teachers explained to us that they knew bilingualism was an asset, but they were not sure what role it played in the lives of second and third-generation children. They wondered whether Bangla was still necessary in the classroom or whether children were learning sufficiently through English only. They also wondered how bilingual strategies could be used in the mainstream classroom, particularly when some children spoke only English or languages other than Bangla.*

*Kenner et al. (2008:93)*

## 1.1 Statement of the problem

It is generally taken as axiomatic in the Second Language Acquisition (SLA) research community that if multilingual learners are given the opportunity to make what some call ‘judicious’ or ‘systematic’ (e.g. Atkinson 1987, Hammerly 1991, Butzkamm 2003, Lee 2012) use of their first language (L1), this will confer an advantage on the learning of their second language (L2) and on their learning of more general curriculum content. By way of example, consider these expressions of support for such approaches.

“It is not difficult to think of several general advantages of judicious use of the mother tongue.” (Atkinson 1987:242).

“I would estimate that remote SL [second language] instruction that fully takes into account and judiciously uses the NL [native language] of the students can be twice as efficient (i.e. reach the same level of SL proficiency in half the time), without any loss in effectiveness, as instruction that ignores the students’ NL.” (Hammerly 1991:151).

“The mother tongue is, for all school subjects, including foreign language lessons, a child’s strongest ally and should, therefore, be used systematically.” (Butzkamm 2003:30).

Emphatic as these exhortations are, they tend to stop short of fully defining what is meant by ‘judicious’ and ‘systematic’, and offer little more than “vague” (Hall and Cook 2012:292) descriptions of how, and to what ends, L1 might be incorporated into the educational diet of multilingual learners who attend schools where the language of instruction is their L2. For example, Gibbons (2009:92) suggests that teachers should “use the students’ first language to explain the key points of the text [...] prior to having them read it in English”. The Department for Education for England suggests to teachers that “Home language can be used to develop higher-order literacy and cognitive skills [by, for example] discussing characterisation or cause and effect” (Bourne 2002:76). Without clarity over what exactly constitutes ‘use’ in these examples, and without empirical evidence to demonstrate its substantive effects, teachers are left to make pedagogic decisions - which require the investment of time, effort and money - without a clear understanding of what those investments entail and what the results of doing so will be.

Perhaps as a result of this lack of clarity we also often see widespread rejection of the principle of L1-mediated learning among the people who, theoretically at least, stand to benefit from it. Schools with multilingual student bodies, teachers with classes of multilingual pupils, parents of multilingual children and, most importantly, the children themselves all stand to gain by implementing L1-mediated approaches, if such approaches work. However, whether because of formal school policies that outlaw L1 (Jackson 2015), latent reluctance of teachers to countenance its use (Cook 2001), or the wishes of parents who view a strictly English-only environment as the best option for their children in a competitive, Anglophone world (Jacobs 1998), the reality is that use of L1 in L2 classrooms is sometimes viewed with suspicion, and tends to be rejected or ignored as a possible pedagogical tool.

The following examples illustrate this rejection.

“Parents rely on the school to employ all reasonable means to ensure their children leave the school with an excellent command of English. [...] Immersion<sup>1</sup> (as far as is reasonable and possible) is the most effective method of developing language skills.” (Kajonkiet International School 2014).

“I do feel a definite obligation to avoid [the L1] as much as possible.” (Edstrom 2006:280)

---

<sup>1</sup> In this quote, ‘immersion’ is understood as being used in its colloquial sense to mean educational programmes in which the language of instruction is *only* English.

“Il est donc essentiel que la seule langue de communication dans la salle de classe soit le français. [It is therefore essential that the only language of communication in the classroom is French]” (Nova Scotia Department of Education 2006:14, author’s translation)

Even in cases where a ‘zero-tolerance’ approach to children’s L1s is not adopted by schools, we see a lack of clarity over what form L1 provision should take and what effects it is expected to have. In a series of Institute of Education (IOE) case studies summarising the nature of provision for students learning English as an Additional Language (EAL) in the UK, Wallace et al. (2009) report witnessing celebrations of linguistic diversity, use of bilingual teaching assistants, and same-language ‘buddy’ systems for EAL learners in some schools. In commenting on the effectiveness of these features, however, they reported “little evidence [of] use of first languages as *resources for learning*” (Wallace et al. 2009:10, emphasis added). While the purpose of their report was not to recommend specific teaching approaches, it is interesting that Wallace and colleagues do not define further what they mean by that final clause. Similarly, a more recent report of the language development and school achievement of EAL pupils in English schools (Evans et al. 2016) voices support for the idea that using L1 is beneficial for multilingual learners, but uses only the vaguest terms to describe how this might be operationalised in the classroom, referring favourably merely to “tuition” in the L1, for example.

The quote at the top of this chapter captures the uncertainty around L1 use as it relates to teachers in the English school system. Teachers in linguistically diverse classes in the UK have been given the impression by the SLA research community that there is a causal relationship between use of the L1 and linguistic and academic success for the multilingual learners in their classes. However, the SLA community has not been sufficiently clear about what such use entails, whether its purported positive effects apply to all multilingual learners or just a subset of them, and what the expected outcomes are. This uncertainty needs addressing if teachers are to have confidence in the efficacy of the pedagogical approaches they adopt.

To help address the lack of a clear articulation of ‘whether’, ‘by which methods’ and ‘to what ends’ L1-mediated teaching and learning contributes positively to the education of EAL learners, and to help inform a resolution to the collective uncertainty surrounding L1 use in L2 classrooms, it is reasonable to assume that relevant empirical research must exist. Research that compares clearly defined L1-mediated teaching and learning approaches with alternative approaches which use only students’ L2 would allow teachers to make informed decisions about language use in their classrooms. Even in reports whose main purpose is to advocate for L1-mediated approaches, however, references to empirical research of this kind are notable by their absence. Despite

exhortations from the SLA community therefore, the explicit or implicit rejection of L1 as a pedagogical tool by some schools and teachers may simply be explained by the application of Hitchens' Razor: "what can be asserted without evidence can also be dismissed without evidence" (Hitchens 2003: no page).

## 1.2 Why is the issue of L1 use among children who use English as an additional language an issue that deserves attention?

Children who use more than one language are a sizable minority of the student body in England. In the English school system, this group is referred to formally as children for whom English is an additional language (EAL). The number of EAL learners in English schools has steadily risen for decades. In the latter part of the 1990s the proportion of pupils in state-funded schools in England who were classified as EAL was just under eight percent; 7.8% of the primary school population and 7.3% of the secondary school population (NALDIC 2013). This proportion rose steadily over the following twenty years, so that the aggregate figure in 2018 was approximately 19%; 21.2% in primary schools and 16.6% in secondary schools (DfE 2018a). This is one out of every five children in compulsory education in England, or about six children in every class of 30.

Of course, distribution of EAL learners across schools and regions in England is not uniform, but the bottom-line statistic above serves to illustrate that teaching EAL learners is not a niche specialism. It is not a demographic anomaly, interesting only to an invested few. Schools counting EAL learners among their number are not unusual. Teaching children for whom English is an additional language is, therefore, a living, breathing part of being a teacher in England. Teachers must be equipped with knowledge and understanding that empowers them to meet the educational needs of this sizeable constituent of their student body. The educational experience for some EAL pupils will be profoundly influenced by how well their teachers understand their educational needs and how well they meet those needs. If, as has been suggested, incorporating the L1s of these pupils into English mediated education is likely to help them realise their potential at school, all teachers and their EAL pupils stand to benefit from understanding how this can be done.

## 1.3 Who are EAL Pupils?

Before looking further at factors that contribute to teachers' understanding of the potential L1 has for EAL pupils, it is important to acknowledge the scope of the term EAL as a defining characteristic of this group of children. The Department for Education (DfE) describes the designation 'EAL' as follows:

“A pupil is recorded to have English as an additional language if they are exposed to a language at home that is known or believed to be other than English. This measure is not a measure of English language proficiency or a good proxy for recent immigration.” DfE (2017a:10).

The final sentence in this quote includes two important caveats on the limits to how these data can be interpreted. The first relates to the relationship between EAL and English proficiency. School census data (DfE 2017b) tells us that enormous variation in English proficiency exists among pupils classified as EAL. This ranges from ‘New to English’ (5.3% of all pupils classified as EAL in 2017), through ‘Early Acquisition (10.5%), Developing Competence (19.4%), Competent (22.8%), to the largest group of EAL pupils, those who are ‘Fluent’ in English (33.4%). A recent analysis of English proficiency and attainment in national examinations in one area of London (Demie 2018) found that pupils classified as EAL who had also been assessed as ‘Competent’ or ‘Fluent’ in English outperformed EAL pupils with lower levels of English proficiency. ‘Competent’ and ‘Fluent’ EAL pupils also outperformed their monolingual English-speaking peers by some margin. As for immigration status, there is no *prima facie* reason to assume that there will be a strong association between these data and the classification of EAL. Among recent migrants to the UK are monolingual English children (DfE 2017b) and children whose families first moved to the UK many generations ago. In addition, some pupils classified as EAL have no history of migration in their families; heritage speakers of Welsh and Gaelic, for example.

In addition to English proficiency and countries of birth, there are many other potentially important individual characteristics that are masked by the single designation ‘EAL’. As one might expect, EAL pupils represent enormous diversity in terms of these characteristics. Ethnicity is one such characteristic, with important associations with success at school. Strand, Malmberg and Hall (2015) analysed national exam results by ethnic group and found that ethnicity was a far more important predictor of attainment in these exams than the designation EAL. For example, average attainment in the statutory Standard Assessment Tests (SATs) taken at the end of primary education in England was much higher in pupils classified as ethnically Chinese (both EAL and non-EAL) than the average attainment for all pupils. At the other end of the scale, SATs attainment among pupils classified as Black Caribbean were lower than the national average (again, irrespective of whether they were also classified as having EAL).

There are other characteristics that play important roles in predicting success for EAL pupils, such as length of time in the English education system, prior educational experiences, and socio-economic status. Most immediately relevant to this thesis, however, relates to the part played by the non-

English languages that EAL pupils are exposed to in the home, languages that I have called ‘first languages’ or L1s. Different L1s of EAL pupils have been shown to be associated with different levels of academic success. Demie, McDonald and Hau (2016) examined relationships between attainment in GCSEs (the exams taken at the end of compulsory education in the UK, when pupils are 16) and the L1s of the students who took them. When they disaggregated average GCSE scores by L1, Demie and colleagues found that the highest attaining pupils were those who spoke Marathi and Telugu (both Indian languages). These pupils performed about 30 percentage points higher than the national average at the time. Czech speakers, on the other hand, scored on average 46.6 percentage points lower than the national average.

In addition to the variety of L1s used in the pupils’ homes, levels of proficiency in those languages are likely to vary enormously also. Just because an EAL pupil is exposed to a language other than English in the home, it cannot be assumed that they understand or speak it. Some pupils classified as EAL are exposed to an L1 via their parents or grandparents, but do not use it themselves at all. Some EAL pupils’ use of L1 will be limited to social domains, such as speaking with family and friends. These are very different to the academic language domains so crucial for success in school. Other EAL pupils will have spent some time in schools overseas, or may go to community schools in the UK to maintain and develop their L1 literacy. As a result, these pupils may have very well-developed academic proficiency in their L1.

Unfortunately, the causal mechanisms at play in the relationships between L1 and academic success identified by Demie, McDonald and Hau (2016) are opaque. That is, we are unclear why the Marathi speakers tended to do so much better in their GCSEs than the Czech speakers. Presumably some of the differences can be explained by factors that coincide with language type, such as socio-economic status or culturally mediated attitudes to education, rather than the L1 itself.

All of this is to draw attention to the caution necessary when considering the role that EAL pupils’ L1s might play in supporting their academic and linguistic development. The potential role for L1 among students classified as EAL is likely to be informed by differences in all the factors I have described above, and many more that I either haven’t thought of, don’t know about, or can’t know about. Any statement about what is and what is not good for pupils classified as EAL must take into account the enormous diversity represented by that classification.

## 1.4 Aim of the thesis

Taking into account common (and often unreconciled) assumptions about the place and role of EAL learners’ L1 in mainstream English classrooms, the aim of this thesis is to address vagueness and

uncertainty around L1-mediated teaching and learning approaches in these contexts. The thesis' original contribution to the field is twofold. First, a systematic review of the literature collects and synthesises existing empirical experimental evidence on the use of L1-mediated pedagogical strategies. This demonstrates the extent of what can be claimed, on the basis of experimental evidence, about the use of L1 as a pedagogical tool. The results of the systematic review then inform the second original contribution. This is an unbiased comparison of alternative ways to teach EAL learners English vocabulary, one way that is mediated by the L1 of the learner and one that is conducted in English only. The results of both original contributions will be discussed in terms of their implications for practice and research, and theory.

The thesis is structured in the following way. It begins with a narrative overview of the literature that informs our thinking around the use of L1 in language and content learning. The second part of the thesis is the systematic review. As a piece of primary research in its own right, the systematic review is presented using the standard reporting structure of introduction, methods, results and discussion. The findings of the systematic review inform the nature of the intervention study that constitutes the third main section of the thesis. This too is presented using the standard structure. The thesis will conclude with a general discussion, with implications for practice, theory and future research.



# CHAPTER 2

## LITERATURE REVIEW

### 2.1 Introduction

In this chapter I review the literature that tends to inform the 'orthodox' position about the use of L1 in language learning classrooms. In so doing I will identify areas where this literature delivers concrete empirical evidence to inform teacher practice, and areas where such evidence is found lacking. I will then suggest what is needed to build on this literature to help teachers in linguistically diverse L2 schools to become better informed about the approaches to, and effects of, L1 use in their classrooms.

### 2.2 Overview of relevant research

Much has been theorised about how the brain handles more than one language, and arguments have been made that a multilingual language user builds competence in one language on the foundations laid by the other. Observational research has provided support for this theory, revealing positive relationships between students' L1 and L2 proficiency. We know, for example, that children who are good readers in their L1 are more likely to be good readers in their L2 as well (*e.g.* Chuang, Joshi and Dixon, 2011). As well as correlational evidence we have some evidence from experiments that helps illuminate the causal relationships between L1 and L2 proficiency. For example, research on bilingual education tells us that children who go to bilingual schools tend to do better linguistically and academically than their peers who go to monolingual schools (Krashen and McField 2005, Reljić, Ferring and Martin 2015, Steele et al. 2017a).

We are also informed about the potential pedagogic role for L1 by research that examines the purposes to which L1 is put by multilingual pupils when they engage in educational activities. When given the opportunity to use their L1, multilingual students use it in ways that seem likely to support their learning. For example, they use it for planning, reviewing and keeping themselves on task (Swain and Lapkin 2000a, Duarte 2016). In addition, research into the relationships between L1 use and wellbeing indicates that being given opportunities to use L1 positively affect students' school experiences (Parke et al. 2002, Kenner et al. 2008).

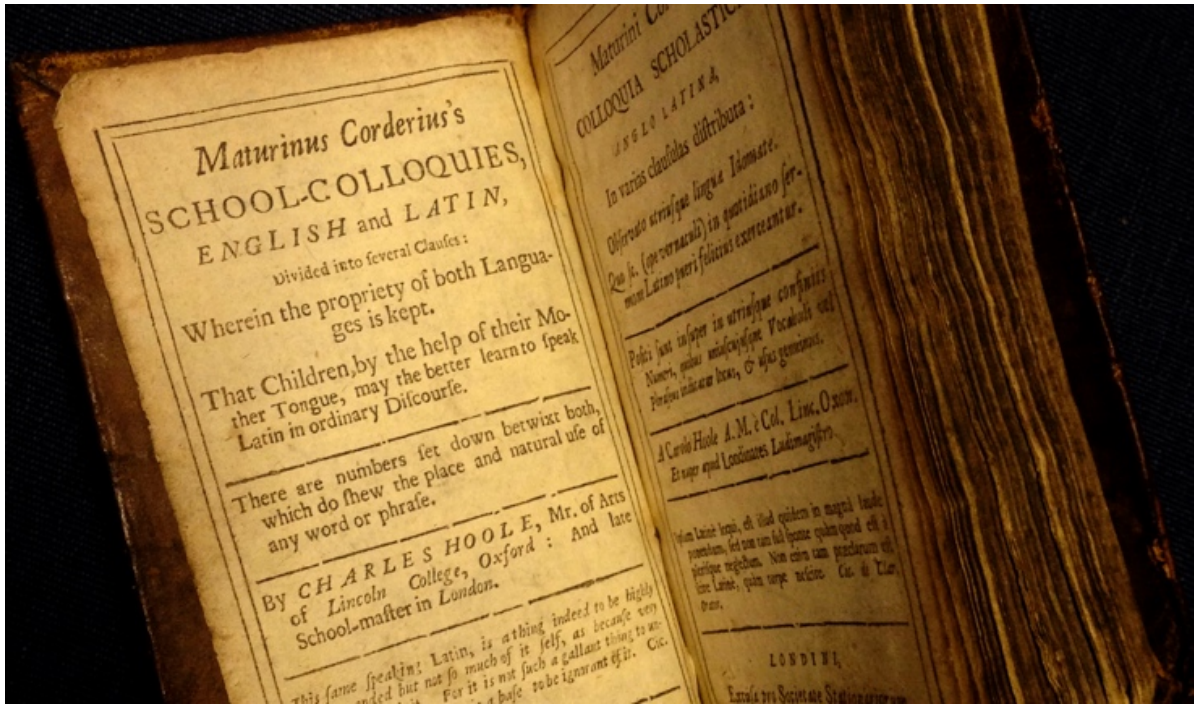
Most of our understanding about the potential role of the L1 has been obtained in contexts with less L1 diversity than is typical in English schools. The volume of research conducted in linguistically

diverse contexts, such as English schools or international schools, appears to be vanishingly small. This is not to suggest that the research that I will summarise in this review is of little value to the English school context, just that it is important to bear in mind the linguistically and culturally diverse character of English schools and how this differs from the contexts in which much of the ensuing research was conducted.

To return to the question posed in Section 1.1: 'Why is the issue of L1 use among children who use English as an additional language an issue that deserves attention?'. Both received wisdom and empirical evidence suggests that systematically and judiciously incorporating multilingual students' L1 into their education might be advantageous. However, research that clearly indicates what is meant by words like 'systematic' and 'judicious' appears to be scarce. Research that clearly demonstrates what the substantive effects of being systematic and judicious are is scarcer still. Moreover, such evidence that exists tends to be derived from contexts that differ in important ways to the English context, most notably in terms of L1 diversity. For example, a large amount of research on bilingual education has been conducted in the USA and Canada where the L1 tends to be shared by all of the pupils, and where teachers are themselves bilingual. This is rarely the case in England. Encouraging teachers to use L1-mediated teaching strategies, without also providing clear descriptions of these strategies nor empirical evidence of their effects, risks teachers and schools investing in resources that may make no difference to pupils outcomes, or may even be counterproductive. Deeper understanding of what L1-mediated strategies are, who they are intended for, and what happens as a result of using them in linguistically diverse schools is vital if teachers in England are to be confident that operationalising them will be worthwhile.

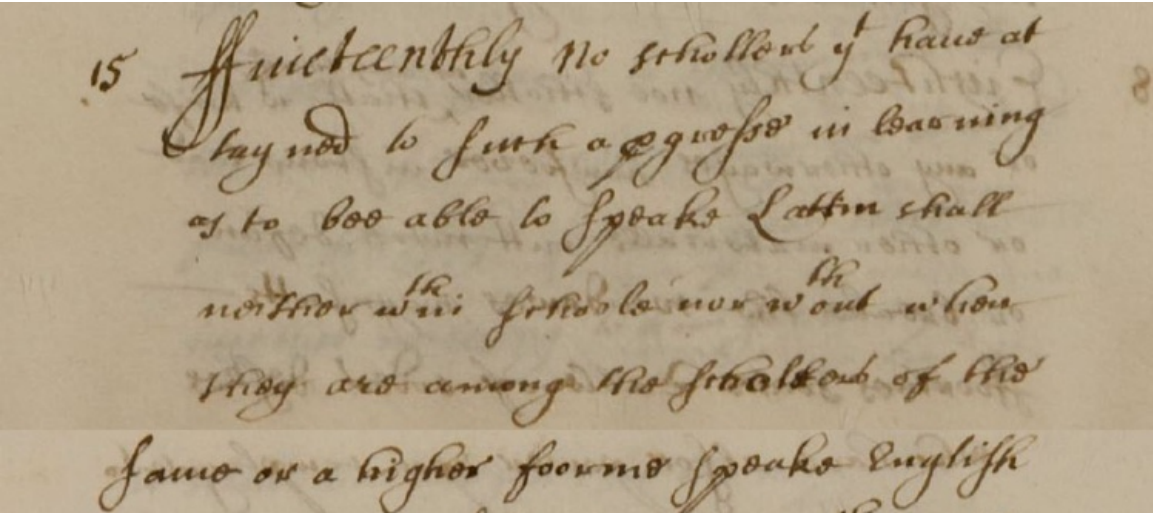
## 2.3 A bit of history

As described in the introduction to this thesis, the position held by specialists in second language acquisition and the position held by schools in which multilingual pupils are educated are often diametrically opposed. Language specialists recognise that a positive relationship between L1 and L2 exists, so they counsel teachers to make judicious use of their pupils' L1s in the service of teaching English. By contrast, many schools tend to assume that allowing multilingual language usage will detract from learning English, confuse the learner, or obstruct effective teaching. Schools thus tend either to prohibit, minimise or simply ignore their pupils' L1s. We saw recent examples of this polarisation in the introduction. However, the argument has run for a lot longer than one might assume. In 1684 Latin teacher and celebrated educationalist Charles Hoole prepared a book of Latin colloquies (short conversations designed to illustrate Latin in ordinary usage) in which he presented the Latin side-by-side with English translations. The title of this book can be seen in the Figure 2.1.



**Figure 2.1.** "...children by the help of their mother tongue may the better learn to speak Latin in ordinary discourse." Maturin and Hoole (1684).

Clearly Hoole felt that his pupils' L1 was an important aide to their learning of Latin. However, consider the second picture (Figure 2.2) taken from the 1656 founding statutes of Adam's Free Grammar school, a Shropshire school in which the language of instruction was Latin.



**Figure 2.2.** Item 15 of the founding statutes of Adam's Free Grammar School, Shropshire.

It states that "No scholars [...] shall neither within the School or without, when they are among the scholars of the same or a higher form, speak English." Adams' Free Grammar School (1656).

These roughly contemporaneous examples of the mismatch between the attitudes of language specialists and the attitudes of school policy-makers towards multilingual pupils' L1s is an almost perfect foreshadowing of the situation three and a half centuries later. In 2002 a DfE guidance document written to inform teachers about the use of L1 in class, says "It is important to make sure bilingual support is used for learning" (Bourne 2002:77). Meanwhile, just as Adam's Free Grammar school was in the 17<sup>th</sup> century, some modern-day school policies could not be more opposed to that advice. Consider this Vancouver high school's attitude to L1 use: "The English language only policy requires that all staff and students speak English inside the school, including the lobby area and the front desk [...] You have come to Pattison High School in order to improve your English. This can only be done if you try to speak English at all times. Your teachers will insist you speak English to them in order to help you develop your English skills and to teach you better" (Pattison High School, 2017).

That such completely opposed views about the place of L1 in the education of multilingual pupils have persisted, essentially unchanged, for more than three centuries is remarkable. Before addressing research that helps to adjudicate the relative merits of these positions, it is worth briefly reviewing the development of the arguments over the 350 years since Hoole said L1 is helpful and Adam's Free Grammar said otherwise.

### 2.3.1 From Grammar Translation to Second Language Submersion

Until the Reform Movement at the end of the nineteenth century, L1 was considered an integral and unremarkable element of learning an L2. In England, the grammar translation method sought to teach the L2 by having learners directly translate back and forth between English and the target language (normally Greek or Latin). This was achieved through memorising words and grammar rules in the context of how the language was used in literature (Knight 2001). By the end of the 19<sup>th</sup> century the language teaching community began to question the effectiveness of this method and proposed new ways of promoting language learning. This involved moving away from the language of literature and towards teaching that concentrated on 'authentic' examples of language use. This approach was pioneered by Maximilian Berlitz whose language schools rely on the so-called Direct Method. In the Direct Method only the target language is permitted. Berlitz's method was extremely influential in informing approaches to language teaching since his first school opened in 1878, and is still very much in evidence in language schools all over the world today.

Although methods like Berlitz's received increasing support over the first half of the twentieth century, and teachers responded by adopting approaches that edged L1 out of the L2 classroom, L1 nonetheless continued to assert influence. Lado's contrastive analysis hypothesis (1957) claimed that by analysing a learner's L1 for similarities and differences between it and the L2, language

teachers could predict areas of difficulty and adapt their teaching accordingly. While this is not the same as supporting learning through pupils' L1, it demonstrates an acknowledgment that L1 is an important mediator of proficiency in the L2. However, perhaps because Lado positioned L1 as responsible for "linguistic distortions" (Lado 1957:1) of the L2, the contrastive analysis hypothesis was used to further arguments that L1 should not be welcomed into the language teaching classroom. Support for the contrastive analysis hypothesis was built on assumptions from the behaviourist school of psychology. This school held that language was developed through a series of learned behaviours (see De Bot, Lowie and Verspoor 2005) and that the job of the language learner was essentially to unlearn what he or she already knew about language (or at least hide it away, lest it interfere). As a result, the use of the L1 continued to be strongly discouraged.

In the 1970s and 1980s the findings of empirical research (e.g. Bailey, Madden and Krashen 1974; Meisel, Clahsen and Peinemann 1981) began to provide *post hoc* support for the views informing the Direct Methods established in earlier decades. This support was characterised by claims that acquisition of L2 followed the same pattern as acquisition of L1, and was learned independently of it. The theory claimed that the vast majority of errors in acquisition could be explained in developmental terms rather than as a result of negative transfer from the L1, as proposed by Lado (Benson 2002). This led to the argument that conditions for L2 learning should be very similar to the conditions under which we learn our L1. This argument was helped by Krashen's (1982) influential *input hypothesis*, which suggests that the sole prerequisite for language acquisition is for learners to receive comprehensible input. Cook (2001a:504) describes "fierce criticisms" of this hypothesis, but not before it had exerted enormous influence on how people thought L2s were learned and therefore the approaches they used to teaching it. This influence led to widespread use of so-called *natural methods* of language teaching, in which the L1 is explicitly unwelcome. Audio-lingualism, The Silent Way, Total Physical Response, and Communicative Language Teaching, are all *natural methods* that enjoyed prominence to a greater or lesser extent in the latter half of the twentieth century and the beginning of the twenty-first (Knight 2001). The prominence of such methods has, in turn, helped to entrench the popularity of submersion<sup>2</sup> models of teaching in mainstream schools. In these it is assumed that multilingual learners will "pick up" the language they need just by being submersed in it (Costley, 2013:278).

---

<sup>2</sup> Submersion is defined by Gallagher (2008:11) as "programmes for bilingual children that actively suppress their first language, culture and identity. (Submersion is the teaching of second-language students solely through the target language [...])" The term is rarely used by schools and teachers who adopt this model. Rather, it tends to be used pejoratively by observers to criticise the practice. Schools that adopt this practice are more likely to self-identify 'immersion' schools. It is important, however, not to confuse this interpretation of the term with its use to describe a specific variety of bilingual education, popularised in Canada, where instruction in the L1 is always present (Murphy 2014).

The insistence that languages are best learned without recourse to the L1 came to be known as the monolingual principle. Commenting in 1984 Howatt (1984:289) says that "the monolingual principle, the unique contribution of the twentieth century to classroom language teaching, remains the bedrock notion from which the others ultimately derive". It is testament to the persistence of this principle that, writing nearly a quarter of a century after Howatt, Cummins noted that "policy and practice operate as though the 'monolingual principle' had been established as axiomatic and essentially 'common sense'" (Cummins, 2007:224). As I will demonstrate, this does not reflect current understanding of how additional languages are learned.

### 2.3.2 The rehabilitation of L1 as a necessary component in effective language learning

Cummins' use of the words 'common sense' suggests why the monolingual principle has so successfully persisted. It seems intuitively right. The logical argument goes: we did a very good job of learning our L1 with no help from any other languages, the best way, therefore, to learn our L2 is to mirror that process. In *The mother tongue in the classroom: a neglected resource?* Atkinson (1987) identified four additional reasons for this 'common sense' interpretation of the process of learning a new language and thus why schools have tended to minimise or ban use of students' L1s. These were: the L1's association with the discredited grammar transition method; a 'backwash effect' of monolingual English teachers working in language schools; the considerable influence of Krashen's work; and the truism that one learns English by speaking English. Atkinson suggests that these influences are unhelpful, stating that "to ignore the mother tongue in a monolingual classroom is almost certainly to teach with less than maximum efficiency" (Atkinson 1987:247). Atkinson was not the first to make this suggestion, and many commentators before and after he wrote this article make similar arguments (*e.g.* Hoole 1984, Hammerly 1991, Butzkamm 2003, 2007, Nation 2003, *inter alia*).

One of the most fundamental, if unconscious, misunderstandings on the part of those whom Atkinson was hoping to convince is the tendency to think of new-to-English students as *tabulae rasae* or 'blank slates' (see Swan 1985a), with no prior knowledge that would help the process of English acquisition. This way of thinking was exemplified in the interim report of the Rochford Review, set up in 2015 by the UK government to review assessment procedures for children working below the standard of the national curriculum tests. The authors of the report characterised EAL learners as children who "may be new arrivals to the country with undeveloped *language skills*"

(Rochford 2015a: 4, emphasis added)<sup>3</sup>. This is false. Second language learners bring a corpus of language – their L1 – to the English language classroom. Using this corpus, they have developed their knowledge of the world, their communicative abilities, and, if they have attended school in their home countries, a wealth of subject knowledge and a familiarity with the associated academic discourse (Swan 1985a). As I will demonstrate, research that has examined the relationships between L1 and L2 skills and knowledge suggests that the wealth of skills and knowledge that multilingual learners bring to the classroom will be extremely important in to the development of their L2. The SLA research community is now widely agreed that L1 plays a positive role in L2 learning (e.g. Harbord 1992, Atkinson 1993, Cook 1999, Butzkhamm 2003, Swain and Lapkin 2000a, Nation 2003, Li Wei and Martin 2009, Gibbons 2009, Alegría de la Colina and del Pilar Garcia Mayo 2009, He 2012, Yiakoumetti 2011, among many others).

In the next section I will summarise research that has influenced how we view multilingual learners' L1s in relation to their L2, and how this view informs the position that L1 supports learning of and through the L2.

## 2.4. Overview of the theoretical position

### 2.4.1 Common Underlying Proficiency

Early attempts to conceptualise the way multiple languages coexist in the mind of a multilingual person tended to rely on the view that the brain is a rigid structure with a finite capacity to store information (see Gallagher 2008). This, in turn, informed the widely held belief that introducing an additional language would result in competition for space between it and the L1; cognitive deficits such as low intelligence would be the result (Legarreta-Marcaida 1981, Cummins 1981). This view is illustrated in a 1953 '*Review of the literature on the effects of bilingualism upon the measurement of intelligence*' (Darcy 1953), in which the author concludes that "The general trend [...] has been toward the conclusion that bilingualists suffer from a language handicap when measured by verbal tests of intelligence," (Darcy 1953:50).

Early interpretations of brain architecture and how it relates to language storage is famously represented by Cummins' (1980) diagram showing a cross section of a human head containing two

---

<sup>3</sup> The authors of the Rochford Review modified their wording in later iterations of the report after this falsehood was pointed out to them. It now states that EAL learners "may be new arrivals to the country with undeveloped *English* language skills" (Rochford 2015b, emphasis added).

balloons. Each balloon represents a set of proficiencies that underlie the linguistic competence associated with each language. The balloons are completely separate, so that as one is inflated (as the proficiencies that underlie that language are added to) it squeezes out space for the other to do likewise. The logical conclusion of this model is that unless one language is allowed to develop fully, and thus to completely flatten the other, neither language will be allowed to develop to its full potential. As a consequence, the person will experience cognitive deficits. Cummins referred to this as a "Separate Underlying Proficiency (SUP) model of bilingual proficiency" (1980:91), and argued that it was a fundamental misrepresentation of how the bilingual mind works.

Cummins (1980) proposed instead that the bilingual brain operates under a *Common Underlying Proficiency (CUP)* model. Here, instead of two mutually isolated language balloons, the proficiencies that underlie linguistic competence are represented by a single balloon. Importantly, the balloon has two necks. One neck represents a channel for linguistic development in the L1 and the other represents a channel for linguistic development in the L2; in particular, development of what Cummins called Cognitive/Academic Language Proficiency, or CALP (1980:83). Developing CALP via either channel, argues Cummins, contributes to a shared, or common, underlying proficiency used by both languages.

## 2.4.2 Basic Interpersonal Communicative Skills and Cognitive Academic Language Proficiency

It is important to understand what Cummins meant by CALP and how it differs from what he called Basic Interpersonal Communicative Skills, or BICS (Cummins 1980:84). BICS is the kind of language that we use in day-to-day interactions, and can be described as interpersonal social language. It is often highly contextualised, relates directly to our personal experiences, and is cognitively undemanding. For children, it is the language of the playground, the lunch hall, and the home. With very few exceptions, everybody becomes proficient in BICS in their L1 whatever their general levels of academic aptitude. There may be some surface differences between individuals' BICS proficiency, such as oral fluency, but these are not meaningfully related to measures of academic performance. That is to say, we cannot reliably predict who will become an engaging conversationalist, say, by looking at other aspects of their cognitive development.

Given this almost universal pattern of apparently straightforward BICS acquisition in L1, there is every reason to assume that this kind of language is developed relatively straightforwardly in the L2, given the right opportunities. Indeed, it usually is (Murphy 2014). The cognitively undemanding, highly contextualised, socially mediated opportunities needed to develop BICS are readily available in typical school settings, and children are often well motivated to take advantage of them.



However, in addition to BICS, the capacity to use a language proficiently, especially in educational settings, relies on Cummins' construct of CALP. CALP comprises the skills needed to manipulate language and make meaning in decontextualised academic tasks. For example, it is CALP that we use when we hypothesise, justify, classify, synthesise, evaluate, and infer. CALP allows children to build conceptual understanding of the relatively abstract phenomena they routinely encounter in school. CALP is strongly related to literacy and other cognitive skills. It is not acquired in the apparently automatic way that BICS is, and must often be taught explicitly. The importance of CALP as it relates to the development of multilingual pupils relies on the understanding that once an element of CALP has been acquired through one language it does not need to be reacquired in the other language. Cummins gives the following illustration:

“Pupils who know how to tell the time in their mother tongue understand the *concept* of telling time. In order to tell time in the second language [...], they do not need to re-learn the concept of telling time; they simply need to acquire new labels or ‘surface structures’ for an intellectual skill they have already learned.” (2001:18, original emphasis).

Cummins called this symbiotic relationship between L1 and L2 "linguistic interdependence" (1979, 2001). Using the terms 'Lx' and 'Ly' to denote the two languages known to a bilingual person, he expressed the nature of this interdependence formally as follows:

“To the extent that instruction in Lx is effective in promoting proficiency in Lx, transfer of this proficiency to Ly will occur provided there is adequate exposure to Ly (either in school or environment) and adequate motivation to learn Ly.” (Cummins 1981:29).

### 2.4.3 Threshold Hypothesis

At this point it is important to acknowledge another of Cummins' hypotheses, which relates to the linguistic interdependence hypothesis. This is the threshold hypothesis (Cummins, 1979). The linguistic interdependence hypothesis suggests that the languages known to a multilingual learner are mutually supportive, and that therefore these learners' existing L1 skills can help them when they begin to learn in a new language. If existing L1 proficiency positively contributes to overall language development, including L2 development, one would expect to see evidence of this relationship in children attending bilingual schools. Bilingual schools provide the adequate exposure and (one hopes) the motivation seen by Cummins as necessary for linguistic interdependence to evolve. Therefore, it is reasonable to conclude that skills already developed in L1 ought to appear relatively rapidly in multilingual students' practices in their other language. In well executed bilingual programmes they do (see Section 2.7 for a summary of research on bilingual education). However, Cummins and others noticed that this wasn't always the case. In some studies of bilingual education,

some children appeared to benefit from being in the programme while others did not. In the latter case, children's linguistic performance in both the L1 and L2 was considered to fall short of native-like proficiency. This led to the coining of the term "double semilingualism" (Paulston 1975:27) to describe incomplete control of both L1 and L2. Paulston translates and quotes Swedish sociolinguist Bengt Loman suggesting why this state of affairs might exist. "Semilingualism", he says "has especially been observed in individuals who have since childhood had contact with two languages without sufficient or adequate training and stimulation in either" (Loman 1974, quoted in Paulston 1975:27).

The concept of semilingualism has received heavy criticism, and is now regarded as, at best, inaccurate and unhelpful (MacSwan 2000, Petrovic and Olmstead 2001). Nonetheless, Cummins observed that some multilingual learners did worse at school than the linguistic interdependence hypothesis would have predicted. Cummins argued that it was the insufficient training and stimulation in the L1 highlighted by Loman that was responsible for the failure of these children to thrive in a bilingual environment. Broadly speaking, children who had had ample opportunity to maintain and develop their L1 alongside their L2 did better than children whose L1 was comparatively poorly supported. For example, a study of children in a Ukrainian-English bilingual programme in Canada (Cummins and Mulcahy 1978) found that children who had plentiful exposure to Ukrainian in their homes demonstrated better meta-linguistic awareness than their peers who had comparatively little exposure to Ukrainian in their homes.

From observations of this kind, Cummins concluded that, for the benefits associated with linguistic interdependence to emerge, proficiency in both languages needed to have reached a certain level; or crossed a theoretical threshold. Once crossed, the mutually supportive relationship between L1 and L2 would convey the expected benefits on the child's learning. If one language or the other did not cross this threshold, linguistic interdependence would be curtailed. In practice this means that children whose L1s are suppressed (through, for example, school policies prohibiting use of the L1, or by moving to a country where their L1 is not widely spoken) are not in a position to take advantage of the benefits that bilingualism makes available to their more proficient peers.

## 2.5 Pedagogical implications of the theory

### 2.5.1 Using the L1 to activate prior subject knowledge

The potential relevance of the linguistic interdependence hypothesis to the focus of this thesis is twofold. The first relates to understanding subject matter and the relative efficiency of each language for retrieving subject knowledge developed through the L1. If, for example, pupils can tell

the time in their L1, clearly there is value in drawing on that understanding when reviewing telling time in the L2. In the pedagogical literature this is commonly referred to as ‘activating prior knowledge’ (e.g. Conteh 2015, Haynes and Zacarian 2010, Gibbons 2009). Here, teachers give their pupils opportunities to retrieve existing elements of their understanding of a topic through, for example, direct questioning, paired discussion, constructing concept maps and so on. They then build on that prior knowledge with new learning. The question of importance is whether using the L1 to activate prior knowledge is more efficient and more productive than using the L2. That is to say, are there meaningful differences in pupil outcomes when teachers use L1 to help draw pupils’ attention to what they already know about a subject, compared with when they use the L2?

### 2.5.2 Using the L1 as a basis for linguistic development

The second implication concerns the development of linguistic competence in the L2. Butzkamm (2007) draws on Swan’s (1985a) invocation of a *tabula rasa* to remind us that children do not come to the English L2 classroom as linguistic blank slates. Indeed, by the time children are three or four years old, that is *before* they start formal schooling, they have usually acquired conversational competence in the language of their homes (Murphy 2014). Older learners who begin instruction in L2 after considerable time spent learning through their L1 at school will have developed beyond the conversational competence of preschoolers. They will usually be competent in the more complex linguistic rules governing communication of abstract concepts in their L1; Cummins’ notion of CALP. Butzkamm illustrates just how far from blank these scholarly slates are when they arrive at the L2 classroom:

“As they grow into their mother tongue (1) they have learnt to conceptualize their world and have fully grasped the symbolic function of language; (2) they have learnt to communicate; (3) they have learnt to use their voice and to speak; (4) they have acquired an intuitive understanding of grammar and have become aware of many of the finer points of language; (5) they have acquired the secondary skills of reading and writing.” (Butzkamm 2007:71).

There is always something upon which teachers can build as they introduce new concepts and new language. A pupil who can handle inflections such as *-ment* and *-eur* in French already knows how to handle the English equivalents *-ly* and *-er*. A pupil who can synthesise phonetic Thai script to form words has a skill that can be immediately applied to decoding the phonetic script of English. The question of importance here is whether strategies based on theoretical understanding of linguistic interdependence facilitate cross linguistic transfer more efficiently and more productively than approaches that are not. That is, will multilingual learners’ command of L2 be improved if we use L1-mediated strategies for learning?

## 2.6 What kinds of research helps us to understand the potential for using the L1 in the L2 classroom?

We have seen that the linguistic interdependence hypothesis (Cummins 1979) predicts that someone's level of proficiency in their L2 is partially a function of their level of proficiency in their L1. That is to say, the better developed the learner's L1 is, the better developed the L2 is likely to become, assuming adequate opportunities to learn the L2 are made available. Conversely, the hypothesis states, that if L1 development is arrested for some reason (by prohibiting its use, say), this will exert a limiting effect on L2 development. Cummins' hypothesis has been extremely influential in informing assumptions about appropriate pedagogy for multilingual learners.

The theoretical underpinnings of this hypothesis can be tested by examining a variety of different types of research. The most relevant types of empirical research address the following areas, each of which I shall explore more fully below.

### ***Research on cross linguistic transfer***

There is a large body of research that reveals relationships between L1 proficiency and L2 proficiency. This research helps us to understand the implications of introducing English at different stages of linguistic development, and of either removing or maintaining L1.

### ***Research on bilingual programmes of education***

This type of research explores the relationships between bilingual schools and the educational success of multilingual learners. There are many different types of bilingual programme, each with its own aims and ideological positions. Knowing how these different types of programme contribute to the educational attainment of multilingual learners helps to suggest ways that the principles of bilingual education can be adopted in non-bilingual programmes.

### ***Research on how multilingual learners use their L1s***

This type of research observes multilingual learners engaging in learning tasks and examines how they use their L1s. This research helps us to generate hypotheses about how teachers can facilitate these processes in the classroom.

### ***Research on teaching strategies that use multilingual learners' L1s***

It is not always possible to provide bilingual education to multilingual learners, especially in contexts in which many different L1s are represented. This type of research explores the effectiveness of 'stand-alone' strategies that use multilingual learners' L1s as tools to support their understanding in L2 contexts. Evaluation of these strategies is perhaps most relevant for contexts where providing L1

input is complicated by the linguistic diversity of the student population. In these contexts, 'stand-alone' strategies are amenable to incorporation into a teacher's pedagogical repertoire on a needs basis.

Each of these research domains has its own assumptions, goals and expectations and is therefore more or less suited to different educational contexts. For example, research into bilingual programmes in the USA may be directly applicable to other countries in which bilingual programmes are viable (for example, countries with sizeable numbers of multilingual learners who all use the same L1), but less so for countries with linguistically diverse school populations. I will summarise research from the fields outlined above and discuss the implications for understanding how L1 can be incorporated into the educational diet of multilingual learners in L2 schools.

### 2.6.1 Research on cross linguistic transfer

Many studies providing evidence that supports the linguistic interdependence hypothesis have examined associations between proficiency in L1 and proficiency in L2. In the main, studies of this kind tend to find positive relationships between L1 and L2 in similar or related language proficiencies. These associations are true for lower order skills, such as phonic knowledge; higher order skills, such as comprehension and fluency; for languages with similar orthographies, for languages with different orthographies, and have identified transfer both from L1 to L2 and from L2 to L1.

### 2.6.2 Transfer of lower order skills

In studies of languages that use alphabetic script, L1 word decoding, spelling, reading comprehension, phonological awareness, receptive vocabulary and listening comprehension have all been found to be positively associated with similar proficiencies in L2. For example, in a study of phonemic awareness (the ability to perceive and manipulate the smallest units of sound in a language, phonemes) among 68 kindergarten children with Spanish L1, Atwill et al. (2007) found strong positive, statistically significant correlations between phonemic awareness in Spanish and phonemic awareness in English. Atwill and colleagues extended their analysis of these data to determine whether this association was related to L1 proficiency. They grouped participants by measures of L1 receptive vocabulary, creating an 'average' L1 proficiency group and a 'below average' L1 proficiency group. In their analyses they detected a strong positive and statistically significant relationship between L1 receptive vocabulary and L2 phonemic awareness in children in the 'average' group, but did not detect evidence of this relationship in the 'below average' group. Thus, in addition to supporting the linguistic interdependence hypothesis, Atwill et al.'s findings also provide support for Cummins' threshold hypothesis.

While Atwill et al. conducted a cross-sectional study comparing L1 and L2 proficiency at one point in time, Sparks et al. (2008) added to our understanding of relationships between L1 and L2 proficiency in a longitudinal study that tracked pupil outcomes over a ten-year period. They collected data on the L1 competence of 54 English-L1 students in their first five years of schooling, then compared this to their L2 proficiency when they were in Grade 10, after two years of modern foreign language study (either French, German or Spanish). They found three patterns. 1) The students' English decoding proficiency (ability to identify a sound from the letter or group of letters representing it) in early elementary school predicted their L2 decoding proficiency in Grade 10; 2) L1 spelling proficiency and phonological awareness predicted L2 spelling proficiency; and 3) L1 reading comprehension predicted L2 reading comprehension. All of these associations were statistically significant and underscore the authors' conclusion that "even several years after students learn to read and spell in their L1, word decoding, spelling and reading comprehension skills transfer from L1 to L2" (Sparks et al. 2008:162). They also point to the pedagogical implications of these findings regarding the importance of a well-developed L1 in supporting later L2 acquisition.

The direction of cross linguistic transfer is not restricted to skills in L1 transferring to the L2. Murphy et al. (2014) conducted a randomised trial to determine whether learning an L2 has a facilitative effect on learners' L1 literacy proficiency. One hundred and twenty children aged between seven and nine years were randomly allocated to one of three comparison groups. Two groups took part in a fifteen-week foreign language programme, in either French or Italian. The third group continued with their usual learning. The foreign language programmes focused on the teaching of vocabulary using a range of activities designed to draw attention to the sounds and spelling of target words. Following the intervention, they found that both foreign language groups outperformed children in the control group on tests of English reading accuracy and phonological processing. This difference was greater for the group that had studied Italian. Murphy and colleagues discuss possible aetiologies for the observed differences and suggest that close study of a new orthography heightened the intervention groups' "appreciation for, and understanding of language as a system of sounds, words, and structure that can be manipulated in different ways." (Murphy et al. 2014:19). Based on this study, it appears that meta-linguistic awareness of this sort applies not solely to the language in which it was developed, nor does transfer occur only from the stronger language to the weaker one. Rather, metalinguistic awareness informs understanding of all the languages known to a multilingual learner.

The studies summarised in this section address lower order literacy skills associated with phonetic awareness, and they largely confirm the reported findings of similar studies (Comeau, Cormier, Grandmaison and Lacroix 1999; Cummins 1998; Deacon, Wade-Woolley and Kirby 2007;

MacCoubrey, Wade-Woolley, Klinger and Kirby 2004; Tingley et al. 2004; *inter alia*). Other studies have explored the cross-language relationships between higher order literacy skill, and will be discussed in the next section.

### 2.6.3 Transfer of higher order skills

Much like the findings of studies of lower order skills, studies that have evaluated associations between L1 proficiency and L2 proficiency in higher order skills, such as comprehension, vocabulary knowledge, and grammatical awareness, have tended to report positive associations too. In a study of higher order reading skills associated with comprehension, van Gelderen et al. (2007) followed 389 adolescent Dutch L1 learners of English from Grade 8 to Grade 10. They took measures of participants' reading comprehension in both languages at three points in time over this period. Their analysis revealed a strong positive, statistically significant association between Dutch reading comprehension and English reading comprehension; an association that strengthened over time. In addition to familiarity with the phonetic alphabetic script used in Dutch, and its application to reading in English, the authors suggest that learners also apply skills developed through their L1 relating to vocabulary and grammar. They note that these associations were there from the start of the students' L2 instruction; there was no lag time between the development of L1 skills and their application in L2. This observation lends support for the authors' suggestion that higher order skills such as knowledge about text structures and genres are shared between L1 and L2. Or, as the authors conclude, "Good L1 readers are, for the most part, also good L2 readers" (van Gelderen et al. 2007:487).

A study of German elementary school children learning English (Gebauer, Zaunbauer, and Möller, 2013), reported moderate positive correlations between the L1 and L2 in reading comprehension and reading fluency, the latter of which is thought to support reading comprehension in the early grades (Jenkins et al. 2003). Two-hundred and twenty German L1 students in Grades 3 and 4 (eight to ten years old) of 'partial English immersion programmes' in German elementary schools took part in the study. The participants were allocated to the immersion programme at the request of their parents and teachers, a potentially important confound, which I will return to later. They were taught in English in all areas of the curriculum from Grade 1, with the exception of German Language Arts. The study explored the hypotheses that L1 and L2 reading proficiency in Grade 3 positively predicts L1 and L2 reading proficiency in Grade 4. Their statistical analysis found moderate, positive, statistically significant correlations between L1 and L2 on all measures of fluency and comprehension. Unlike Murphy et al.'s (2014) study, this study was not designed to identify casual relationships. As with all correlational research, therefore, one cannot say with any certainty what is responsible for the associations they observed. Does L1 support L2, does L2 support L1, or is a third factor

responsible for development in both? Indeed, the authors identify this limitation to their research, noting that a study by Kuska, Zaunbauer and Möller (2010:143) found that children in immersion programmes had “better learning and memory performances (sic)” and higher socioeconomic status than their peers in monolingual programmes. The fairly closely related proficiencies in German and English among participants in Gebauer et al.’s study could be a function of these factors. Being good readers, as van Gelderen et al. put it, might be a characteristic of children who are selected to attend bilingual programmes (recall that participants had been chosen to attend the programmes by their parents and teachers). Therefore, the closely related reading proficiencies in L1 and L2 found by Gebauer et al. might be merely a function of that, rather than a result of bilingual education *per se*. Nonetheless, these findings are indicative that a relationship exists between these higher order literacy skills across languages, whatever their antecedents.

#### 2.6.4 Transfer across dissimilar languages

The studies described above demonstrate evidence for cross linguistic transfer of higher order literacy proficiencies between orthographically similar languages (languages that use similar systems of writing, in this case Roman script). In addition, there is evidence that proficiency in one language is positively related to proficiency in another language even when orthographies are very different. English uses a phonetic script in its writing system. Mandarin Chinese uses pictographs. The way in which these two systems are decoded by a reader to form words is different. In phonetic script, individual letters represent different sounds. These sounds are combined to form a word. Because of this, competent readers of phonetic scripts like English can correctly say a word that they encounter in a text without ever having heard it before. In pictographic scripts like Chinese, each character represents an entire morpheme (the smallest unit of sound that carries meaning). As such, the sound and meaning of the morpheme depicted by each pictograph must be memorised in order that a word may be read. Words that the reader is unfamiliar with cannot reliably be sounded out in the way that phonetic orthographies make possible. As a result, one might expect any relationship between orthographic skills in the two languages to be weak, if it exists at all. Indeed, some have argued that if Mandarin Chinese readers try to apply the lower order orthographic skills associated with reading Mandarin Chinese to learning to read English it is likely to be counterproductive (Keung and Ho 2009). However, a study in Taiwan (Chuang, Joshi and Dixon 2011) examined educational data for a random sample of 30,000 Mandarin Chinese-L1 ninth-grade students studying English as a foreign language and found a strong positive statistically significant correlation between reading proficiency in Mandarin and in English. About 63% of the variation in English reading scores could be explained by Mandarin reading scores. This association held across demographic groups, such as rural students, urban students, males, females, and for different lengths of time spent learning



English. It also held across grammatical knowledge, vocabulary knowledge, and reading comprehension in both languages. The authors conclude that cross linguistic transfer of reading proficiency occurs between languages that are structurally very different as much as it does for structurally similar languages. While the research on cross-linguistic transfer between orthographically different languages that preceded this study has reported mixed results, some saying that only certain features of literacy proficiency transfer across dissimilar orthographies while others do not (e.g. Bialystok, Luk, and Kwan 2005; Bialystok, McBride-Chang and Luk 2005; Wang et al. 2005), the size and unbiased way in which data were selected in this study lend authority to the authors' conclusion.

### 2.6.5 Long-term effects attributed to cross-linguistic transfer

Finally on cross-linguistic transfer, extensive, long term research on multilingual students in the USA has identified positive relationships between L1 and ultimate L2 attainment. Thompson (2015) analysed data collected over a nine-year period on more than 200,000 children classified as Limited English Proficient (LEP)<sup>4</sup> in the Los Angeles Unified School District, in California, USA. The key aim of her study was to determine how long it takes for LEP children in the district to be reclassified as English proficient, and to identify whether demographic characteristics in the population of LEP students are meaningfully associated with differences in that time. On entry to Kindergarten LEP children were tested for academic language proficiency in their L1 (mainly Spanish). Subsequently, the children's scores on English language assessments were collected every year until they had reached a level of proficiency in English deemed sufficient to no longer require classification as LEP. Children who entered kindergarten with high levels of academic L1 proficiency were 12% more likely to be reclassified as English proficient by the time they reached Grade 9 than students who entered with beginning levels of L1 academic proficiency. Thompson points out that "these findings are descriptive rather than causal" (2015:29) but that they are consistent with other research that suggests that high-quality pre-school provision in the L1 of LEP children is likely to be of benefit in terms of later attainment in English.

### 2.6.6 Summary of research on cross-linguistic transfer

I have summarised only a handful of the very many studies that have assessed relationships between L1 proficiency and L2 proficiency to assert support for the central tenets of Cummins' linguistic interdependence hypothesis, common underlying proficiency theory, and threshold

---

<sup>4</sup> LEP is a term commonly used in the USA that is only partially comparable to the UK designation EAL. The main difference is that a child in the USA can be re-classified as EP (English Proficient) once they have reached a certain level of mastery over the language. The term EAL makes no such distinction.

hypothesis. The findings of these studies (and many more like them) are used to argue for the importance of good L1 support for children learning an additional language, especially in the early grades.

With the exception of the study by Murphy et al. (2014) all of the studies I have reported here are correlational. They cannot (and, in most cases, do not) assert a causal relationship between teaching that focuses on developing students' L1 and their subsequent attainment in L2. Therefore, while they tell us that children with strong L1s are likely to also have strong L2s they cannot by themselves be used to argue that teaching L1 will improve L2. To make that argument one needs to look at experimental research that compares the effects on L2 outcomes of teaching approaches that use the L1 and teaching methods that do not use the L1. One of the largest bodies of research of this sort assesses the effects of bilingual schooling. This research will be the focus of the next section.

## 2.7 Evidence from bilingual programmes of education

Some of the clearest evidence on the effects of using multilingual learners' L1 in combination with their L2 is derived from a rich and varied literature on bilingual programmes. The preponderance of studies on the effects of bilingual education is thanks in no small part to the 1974 amendment to the Bilingual Education Act in the USA. This Act encouraged individual states to provide bilingual education for language minority children who wanted it. As well as a good deal of resistance from some quarters (*e.g.* Epstein 1977), the Act generated considerable interest in developing bilingual programmes and thus the opportunity to conduct empirical research through which to evaluate their effects. The findings of many of these studies have been combined in a number of systematic reviews and meta-analyses, which I will summarise later.

### 2.7.1 The St. Lambert Experiment: early evidence for the effects of bilingual schooling

By way of illustration of the general trend, I will begin by summarising a study regarded by many as seminal in framing our understanding of the implications of dual language, or bilingual, education. In the 1960s, in Montréal, Canada, a group of English-speaking parents living in a middle-class English speaking suburb of the predominantly Francophone city were concerned that their children would find it difficult to integrate with their French speaking peers if they were unable to speak French well (Lambert and Macnamara, 1969). They felt that the methods used to teach French in typical English language schools fell short of what was needed to help their children become fluent French speakers. As an alternative, these parents were attracted to the idea of French immersion schooling. Immersion programmes vary, but the characteristic shared by them all is that a significant proportion of the school day is taught in the target language, while L1 development is maintained and

developed. This is done, at minimum, by providing regular L1 Language Arts instruction, and often by sharing the medium of instruction of other curriculum subjects across languages. Ultimately most models of immersion schooling aim to use both languages in roughly equal measure after an initial period where L2 instruction dominates (Murphy, 2014). In-keeping with common concerns at the time, the parents worried that enrolling their children in French immersion classes might come at a cost to their English proficiency and possibly their general intelligence. Nonetheless, motivated by a desire to base decisions about their children's education on evidence, the group approached researchers at McGill University in Montréal and worked with them to set up an experimental French immersion programme to assess the effects of bilingual instruction on the things that mattered to them. In their novel programme instruction was nearly all in French in Kindergarten, but with regular English language arts classes. Gradually the proportion of teaching in English increased, so that by Grade 5 half of the instruction was in French and half was in English. The students' proficiency in English language, French language, mathematics, science, and general intelligence was closely monitored for the duration of the project and compared with similar children in monolingual schools, both French-only and English-only.

The St. Lambert Experiment, as this study became known (Lambert and Tucker, 1972), found that the French language skills of students' in the immersion programme developed very well, allowing them "to read, write, comprehend, and speak French with fluency and naturalness" (Lambert, Tucker and d'Anglejan, 1973:158). By contrast, their anglophone peers in schools that took a traditional approach to teaching French did not do so well. In addition to improved French proficiency, after a short period of catch-up following the introduction of English as a medium of instruction for larger proportions of the curriculum, children in the immersion programme developed English skills that were as good as those of their peers in monolingual English schools. They also found that development in science, maths, and general intelligence was similar. In some cases it was better (Lambert, Tucker and d'Anglejan, 1973).

The bottom line of the St Lambert Experiment was that this model of bilingual schooling did not have a detrimental effect on curriculum learning, L1 development, or general intelligence. It also proved to be a more effective way of learning French than the typical approaches used at the time in other types of school.

The St Lambert experiment is the vehicle for the findings of just one experiment in one educational context. It is important to remember that the children involved in the St Lambert Experiment were from relatively affluent, aspirational families who clearly valued multilingualism. The children were well supported outside school in their L1 (English), and English was a prestigious language valued

beyond the immediate community from which the children in the St Lambert Experiment were drawn. The context in which the St Lambert Experiment took place is likely to have had important implications for the success of the programme. None of these factors can be assumed for children who speak minority languages in the UK. To get a fuller picture of the effects of bilingual education for minority language users it is crucial to look at research on bilingual education conducted in a variety of contexts and to draw lessons from the evidence taken as a whole. What follows is a summary of three systematic reviews of research on bilingual education that help us to understand what the effects of using both L1 and English in those contexts are likely to be.

### 2.7.2 Systematic reviews of research in bilingual programmes of education

A number of systematic reviews of research have been conducted to evaluate the effects of bilingual programmes on educational outcomes in multilingual students. Krashen and McField (2005) identified five such reviews, in which the findings of individual studies were synthesised using meta-analyses to arrive at an overall estimate of the effect of attending a bilingual school compared with attending a monolingual school. They conducted what they called a “*meta-meta-analysis*” (Krashen and McField 2005:8) by combining the findings of all five reviews to calculate a summary finding of all of them considered together. Individually and collectively the reviews came to the same conclusion: there is a small and positive effect associated with attending programmes in which instruction is shared between L1 and L2. So remarkable was the consistency of these findings that Goldenburg (2008:15) was moved to observe that “No other area in educational research [...] can claim five independent [reviews] based on experimental studies - much less five that converge on the same basic finding”.

Given the consistency of the findings of the work synthesised by Krashen and McField (2005) it is unnecessary to go into detail for all five reviews. Instead I will summarise the findings of two particularly robust reviews to illustrate the overall findings.

I will first summarise a review by Slavin and Chueng (2005). As well as being the most recent of the reviews identified by Krashen and McField (2005), Slavin and Cheung used a best evidence approach (Slavin 1986) to help locate, appraise and synthesise relevant studies. This means that they performed an exhaustive search for studies that (a) compared bilingual programmes to English-only approaches, (b) attempted to reduce bias in the ways participants were allocated to comparison conditions, (c) measured outcomes such as test scores, and (d) included only studies for which the duration of the programme was at least one school year. In so doing, they re-analysed studies which had been included in earlier systematic reviews, and incorporated any that met their inclusion

criteria. They also provide quite detailed descriptions of the bilingual programmes included in their review. They have thus compiled the most complete and methodologically robust systematic review of those included in Krashen and McField's (2005) *meta-meta-analysis*. However, Slavin and Cheung's review is limited only to studies of bilingual reading instruction. While reading and literacy directly or indirectly underpin learning in virtually all other areas of education it is sensible to also examine other aspects of bilingual education. I will, therefore, also summarise the review by Rolstad, Mahoney and Glass (2005). The scope of Rolstad, Mahoney and Glass's review was much broader than Slavin and Cheung's. It did not discriminate by methodological rigour, and incorporated studies with a variety of different outcome measures, not just those relating to reading. It thus helps to build a more holistic view of the effects of bilingual education, notwithstanding its less robust approach to methodological rigour.

Perhaps unsurprisingly given motivating principles behind bilingual education in the USA at the time, the outcomes of principal interest in much of the literature on bilingual education relate to proficiency in English. It is quite rare to find the effects on L1 proficiency assessed or reported. However, a small number of the studies in both Slavin and Cheung's review and Rolstad, Mahoney and Glass' review included measures of L1 proficiency. Thanks to these inclusions both reviews provide evidence that bilingual programmes are not just good for target language development and curriculum knowledge, but also for the L1.

The focus of these two reviews is North America. North America provided a rich backdrop to investigating the effects of bilingual education in the seventies and eighties because, as I have noted, the Bilingual Education Act created the climate and opportunity for such research in the USA. Bilingual education in the rest of the world is relatively uncommon (Stanat and Christensen 2006). However, the 1990s saw a backlash against bilingual education in the USA, with the effective outlawing of the approach in a number of states (Ovando 2003). The introduction of the English Acquisition Act of 2002 also privileged rapid transition to English, rather than nurturing the development of bilingualism, and encouraged questionable pedagogical and legal practices (Hillner 2005). As a consequence, the climate and opportunity for more research on bilingual education in the USA dwindled. In addition to addressing only the North American context, therefore, the reviews by Slavin and Cheung and Rolstad, Mahoney and Glass (as well as the three other reviews identified by Krashen and McField (2005)) include little recent research. To address this asymmetry in the literature, I will finish by summarising a review of bilingual programmes in Europe by Reljić, Ferring and Martin (2015), which both broadens the geographical scope and updates our understanding in the light of more recent research. Finally, I will also describe a recent study conducted in Portland, Oregon that leveraged the school district's policy of assigning children by lottery to the district's

oversubscribed bilingual programmes, to create one of the few randomised trials of this model of education.

### 2.7.3 Review of bilingual reading programmes by Slavin and Cheung (2005)

The studies reviewed by Slavin and Cheung (2005) covered four different types of bilingual programme. These were:

- *Paired bilingual programmes* - students are taught in both their L1 and English at different times of the day.
- *Transitional bilingual programmes* – students are initially taught only in their L1, then English is gradually introduced and increased until all instruction is in English.
- *Remedial bilingual interventions* – students at risk of low attainment in English subjects are placed in bilingual ‘catch-up’ programmes.
- *Heritage language programmes* – students who may not be highly proficient in their L1, but who have family or cultural connections to it, are taught literacy in that language.

#### ***Paired bilingual programmes***

In Slavin and Cheung’s review, two studies (Huzar 1973, Plante 1976,) compared early primary paired bilingual programmes with submersion programmes for Puerto-Rican students. The classes in Plante’s study were team-taught by a Spanish speaking teacher and an English-speaking teacher. Foundation subjects were taught in Spanish, and English literacy was gradually introduced over the course of two years. Students were given standardized tests of reading in English and Spanish, and tests of general educational development. Plante found that students in the bilingual programme did better than their peers in submersion programmes on all measures. Huzar’s experiment compared two different approaches to reading instruction. In one, students had two reading lessons per day, one conducted in Spanish and the other conducted in English. In the comparison approach all instruction was in English. At the end of third grade the students in the paired bilingual class had better scores on standardised tests of English and Spanish reading than students in the English-only class.

Two more studies in this review reinforced the findings of Plante (1976) and Huzar (1973). Campeau et al. (1975) prepared a series of reports on a variety of different paired bilingual programmes. Their studies included one in which reading was taught in two-hour sessions that alternated between English and Spanish over the two years from Kindergarten and Grade 1. At the beginning of

kindergarten 90% of the teaching was in Spanish, and focused on Spanish phonetics. By the end of the year teaching was 50% English and 50% Spanish. In Grade 1 there were two lessons in English for every one lesson in Spanish. Students in these paired bilingual programmes did much better than their peers in English-only programmes on standardized tests of English and Spanish reading. Another study by Campeau et al. (1975) compared kindergarten and Grade 1 paired bilingual programmes in which students had one Spanish reading lesson per day and one English reading lesson per day. They found that students in the bilingual programme performed substantially better than similar students who had been taught only in English. A later study of paired bilingual Kindergarten and Grade 1 programmes in four schools over two years reinforced the conclusions of the earlier work. It found that students whose reading instruction time was shared between English and Spanish performed significantly higher in English basic skills tests than students who had been taught only in English (Ramirez et al., 1991).

Not all studies of paired bilingual programmes reviewed by Slavin and Cheung found a clear advantage associated with them (Cohen 1975, J.R. Maldonado 1977, Alvarez 1975). Importantly though, these studies did not find an advantage for English-only programmes either. Children did equally well in both. The obvious difference being that students in the paired bilingual programmes had had the opportunity to develop their L1, an opportunity denied to their peers in the English-only programmes.

### ***Transitional bilingual programmes***

As with other forms of bilingual education, the research reviewed by Slavin and Chueng (2005) finds that transitional programmes are associated with higher attainment for multilingual learners compared with the attainment of their peers in English-only programmes (JA Maldonado, 1994; Campeau et al., 1975; Saldate, Mishra and Medina, 1985).

### ***Comparison of transitional and paired bilingual programmes***

In addition to studies comparing bilingual programmes with English-only programmes, Slavin and Cheung included a seven-year study that compared paired and transitional programmes (Gersten and Woodward 1995). The study found that levels of English proficiency reached by students in each programme type was ultimately quite similar. To begin with, students in the paired-bilingual programme had better scores on tests of basic English skills than students in the transitional programme. This is unsurprising; the nature of transitional programmes meant that in the earlier phase of the comparison students in transitional programmes had yet to receive instruction in English. Over time this difference lessened, such that by Grade 7 the two groups of students had similar levels of English proficiency. It is important to recognise here that, while English attainment

reached comparable levels in both groups, only students in the paired bilingual group had developed academic proficiency in their L1. Long term paired bilingual programmes thus develop a valued skill that is denied to students in transitional programmes.

### ***Remedial bilingual interventions***

Two studies reviewed by Slavin and Cheung examined the effects of remedial bilingual reading interventions in secondary schools. In these programmes Spanish-English secondary school students who were at risk of falling behind in their English reading proficiency attended 'catch-up' programmes delivered in Spanish and English, or in English only (Covey, 1973; Kauffman, 1968). Both studies found that students in the bilingual programmes outperformed their peers in English-only interventions on standardised tests of English reading.

### ***Heritage language programmes***

The two studies of heritage language programmes included in Slavin and Cheung's review evaluated the effects of one-year paired bilingual programmes for Louisiana French heritage children and for Choctaw heritage children, compared to English-only teaching (Morgan, 1971; Doebler and Mardis, 1980, 1981). Morgan (1971) found that children taught in both Louisiana French and English did slightly better in some measures of English reading than children in English-only programmes. Doebler and Mardis' (1980-81) found that students in both programmes did equally well in English. Slavin and Cheung emphasise that these children were already proficient in English at the start of the programme.

### ***Summary of Slavin and Cheung's review of bilingual reading programmes***

Overall, Slavin and Cheung reviewed 17 comparisons of bilingual reading programmes and English-only programmes. In twelve of these, students in the bilingual programmes did better than their peers in English-only programmes on measures of English reading. Five of their reviewed studies concluded that bilingual and English-only instruction were equally effective in developing proficiency in English. Those studies that measured L1 proficiency found that students in bilingual programmes did better on that measure than students in English-only programmes. None of the studies concluded that bilingual education had negative effects on English language or academic attainment more generally.

## 2.7.4 Review of bilingual programmes by Rolstad, Mahoney and Glass (2005)

A second extensive review of research into bilingual education was conducted by Rolstad, Mahoney and Glass (2005). They examined studies that compared bilingual programmes with English-only



programmes for language-minority children in grades K–12. These represented a variety of programme types and outcome measures relating to reading, writing, oral language, mathematics, social studies, and science. They do not go into the detail of the characteristics of these different programmes or the specific details of the 17 studies included in their meta-analysis. When they synthesized the outcomes across all studies, they found that, on average, language-minority children in bilingual programmes did better academically and linguistically than children in English-only programmes.

The authors carried out further analysis to compare different types of bilingual programme (transitional and paired) with English-only programmes. In broad agreement with the research reviewed by Slavin and Cheung (2005), they found that paired bilingual programmes were associated with better linguistic and academic outcomes for their pupils than English-only programmes, while transitional programmes had negligible to negative effects when compared to English-only programmes. Thus, they added more evidence that maintaining and developing L1 in step with English, rather than attempting to transition rapidly to English only, is the version of bilingual programme that brings most educational advantage.

Rolstad, Mahoney and Glass also looked at two studies that assessed the L1 proficiency of participants and which used the L1 to assess attainment in reading, maths, science and social studies. Unsurprisingly, bilingual programmes were found to have a large positive effect on L1 proficiency compared to English-only programmes. This advantage also extended to attainment in other curriculum areas. Compared to students in English-only programmes, students in the bilingual programmes were much better able to demonstrate curriculum understanding in English as well as in their L1.

In line with the conclusions drawn by Slavin and Cheung (2005), Rolstad, Mahoney and Glass conclude that “bilingual education is superior to all-English approaches”, that education programmes “which ban or greatly discouraged the use of the native language for instructional purposes cannot be justified” and that “sound educational policy should permit and even encourage the development and implementation of bilingual education programs” (2005:590,572).

### 2.7.5 Review of bilingual programmes in Europe by Reljić, Ferring and Martin (2015)

Reljić, Ferring and Martin (2015) note that previous reviews have tended to concentrate on older research conducted mainly in North America, leaving us comparatively uninformed about the effects of bilingual programmes in other parts of the world. In their review, they performed an exceptionally thorough search for studies that evaluated European bilingual programmes in comparison to

programmes conducted only in the target language (normally the language of the European state in which the schools were located) and found seven reports that add support to the general findings of research in North America.

The studies Reljić, Ferring and Martin reviewed assessed bilingual programmes that used six different L1s (Urdu, Asturian, Basque, Catalan, Gaelic, and Turkish) and four target languages (English, Spanish, Norwegian, and Dutch). The studies reported on outcomes such as literacy, mathematics, civics, linguistic creativity, and written expression, both in the L1 and in the target languages.

As with previous studies, Reljić, Ferring and Martin's synthesis found a small but statistically significant positive effect on academic and linguistic attainment for students in bilingual schools compared to students in target-language-only schools. They conclude that their review "supports bilingual education in Europe, which specifically includes the home language of language minority children" (Reljić, Ferring and Martin, 2015:29).

### 2.7.6 Randomised trial of district-wide bilingual programmes

To conclude this section, I will describe a recently published evaluation of bilingual programmes, conducted in the USA (Steele et al. 2017a). While this is a single study, and its findings largely confirm those of earlier research, it is valuable to include for a number of important methodological reasons.

The Achilles heel of much comparative research into the effects of bilingual forms of education is selection bias. More often than not in comparisons of bilingual and non-bilingual schools (including most of those in the reviews I have described above) participants or their parents *selected* the type of school they attended. In most cases, because of the role that self-selection might play, there are likely to be systematic (non-random) differences in the average characteristics of children in bilingual programmes compared to those of children in non-bilingual programmes. In many studies of bilingual programmes, therefore, a fundamental assumption relating to the internal validity of comparative research is being violated: like is not being compared with like. Because the two groups are systematically different to each other, differences in outcome between them might be a function of something other than the use of two languages in their education. For example, it is reasonable to assume that parents who choose bilingual education for their children over other forms of education hold attitudes towards bilingualism, their L1, their cultural heritage, and possibly to the education sector as a whole that are not shared by parents who do not choose bilingual education for their children. One cannot rule out these parental attitudes as important mediators of their children's educational success. That is to say, the average family characteristics of language minority children in bilingual schools and language minority children *not* in bilingual schools may predispose them to

achieve differently, irrespective of anything the school does. While researchers can perform statistical techniques to attempt to match participants in comparisons of different school types, this can only be done for known characteristics and characteristics that can be measured. These might include, for example, age, socio-economic status, non-verbal IQ, and so on. However, using these techniques, it is not possible to match participants on potentially important characteristics that are unknown and/or unmeasurable. The only way to ensure that the average of both known and unknown participant characteristics does not differ systematically between groups at baseline is to use an unbiased method of allocation, such as randomisation. Ideally, then, for a better understanding of the putative causal relationships between bilingual programmes and success at school children sampled from the same population ought to be randomly allocated to attend either a bilingual or non-bilingual programme. This poses problems for researchers as there can be legitimate ethical and logistical factors that limit the extent to which participants can be randomly allocated to schools. However, organic conditions and policy norms in the city of Portland, USA, where Steel et al.'s (2017a) study was conducted, offer a relatively rare opportunity to satisfy the need for unbiased comparison groups without compromising ethical or logistical considerations, and therefore to generate a more reliable measure of the effects of each type of programme.

Portland has a long history of providing Dual Language Immersion (DLI) bilingual programmes for its residents. The popularity of these programmes is such that they are routinely oversubscribed. To manage this oversubscription the Portland Public Schools Authority allocates applicants to bilingual programmes using a lottery. The pool of applicants, all of whom share the potentially important characteristic that they are motivated to apply to attend DLI programmes, are assigned by lottery to a place either at a bilingual programme or at one of the district's non-bilingual programmes. The lottery thus ensures that the average characteristics of children drawn from this pool at these two school types differ only as a result of the play of chance. Because this method of allocation is understood, respected and used as a matter of routine in Portland, one of the major ethics stumbling blocks (Gueron 2008) for researchers wishing to use random allocation is negated. Steel and colleagues leveraged this existing method of allocation to compare the outcomes of bilingual programmes with non-bilingual programmes which, in their words, "allowed the study to estimate effects *caused* by access to these programs and not by the unobserved characteristics or preferences of families who chose DLI" (Steel et al. 2017b:1, original emphasis).

The study is also notable for its size, duration and outcome measures. In the studies included in the systematic reviews and meta-analyses described above, sample size rarely exceeded the low hundreds, with half of all studies having fewer than 100 participants. This raises questions over these studies' statistical power to detect meaningful differences between groups, particularly if those

differences are assumed to be small. Steel et al.'s study included 1,946 participants, 864 of whom were allocated to DLI programmes and 1,082 who were allocated to English-only programmes. The duration of the study was four years; suitably long to allow for differential effects to emerge should they exist. In addition, the outcome measures were the usual state-mandated accountability assessments of reading, maths, and science. These assessments were all conducted in English. They also collected data on the frequency with which children assessed as Limited English Proficient on entry were reclassified as English Proficient before the end of the trial. Consequently, they were able to frame the findings of the trial in a 'language' familiar and meaningful to teachers.

The trial found that children in Portland's bilingual programmes outperformed their peers in non-bilingual programmes in reading by between 13 and 22 percent of a standard deviation, or the equivalent of an additional seven to nine months in school. They did not detect a difference in maths or science scores. These findings held for children whose L1 was the partner language and children whose L1 was English. That is, both minority and majority language users benefitted from being in bilingual programmes. They also found that LEP children in the bilingual programmes were more likely to be reclassified to English Proficient by Grade 6 than children in the English-only schools. Finally, they found that students in bilingual programmes reached intermediate levels of proficiency in the partner language compared to novice levels of Spanish proficiency attained by participants in the non-bilingual arm of the trial (Spanish was the modern foreign language studied in the English-only schools). Thus, this methodologically robust randomised trial serves as an apposite restating of the general findings of the body of research in this area, going back at least to the 1970s.

### 2.7.7 Summary of research on bilingual education

Research on bilingual programmes of education almost universally concludes that children from minority language communities who attend them do at least as well as their peers who go to non-bilingual schools, and in many cases they do better. This is both in terms of linguistic outcomes and academic outcomes. That is, children in bilingual schools tend to do better, or no worse, in tests of the target language (in most cases English) and tests of curriculum knowledge than children in monolingual programmes. While in some cases these advantages are short-lived or negligible, in all cases children in bilingual programmes are given the opportunity to develop their L1 in academic ways. As such they enjoy an advantage in a highly desirable skill (bilingualism) that is not made available to children who study in monolingual schools. None of the studies included in this review have concluded that bilingual programmes are damaging to multilingual students' L1, their L2, or their general intelligence. These findings provide a rationale for the suggestion that monolingual schools should find ways to incorporate the L1s of their students into their general education. However, just how monolingual schools should incorporate their students' L1s into their general

education is a question that must be addressed by additional research. Research into bilingual programmes is inevitably contextually specific to those programmes. Moreover, it is important to remember that a significant portion of research on bilingual programmes demonstrates no particular advantage to the general linguistic or academic development of children who attend them (albeit no disadvantage either). This suggests that claims that use of the L1 in very different contexts should be interpreted cautiously. It's possible that the 'advantage' of bilingual schools extends only to the development of bilingualism, rather than nurturing a causal relationship between L1 and L2. That said, to begin to understand how the implications of the findings of research in bilingual schools might be applicable in other contexts we can first turn to research that examines the processes and functions of dual language use in education. This is the focus of the next section.

## 2.8 How do multilingual learners use their L1s at school?

There is a considerable body of research that examines the functions served by multilingual learners' L1 when they engage in educational tasks. In some studies, this is done by asking students to 'think aloud' while they carry out a task independently. In others, the conversations between pairs or groups of students are recorded. In both cases the use of L1 and L2 are analysed to determine the purposes for L1 use.

### 2.8.1 Using the L1 facilitates co-operation and lightens the cognitive load ...

Swain and Lapkin (2000a) gave pairs of Grade 8 English L1 students in a French immersion school in Canada collaborative French language writing tasks. They were told that they could talk to each other in any language they liked while they carried out the task. When they analysed the conversations, they found that the students used their L1 for three main purposes: moving the task along, focussing attention, and interpersonal interaction. Swain and Lapkin argue that use of L1 in this way frees up cognitive resources that can be used to concentrate more fully on the language elements of the task. This is reflected in earlier findings by the two researchers (Swain, 1995, 1999, 2000, Swain and Lapkin, 1998, 2000b) and suggests that L1 scaffolds understanding of tasks, improves information sharing, helps with finding appropriate vocabulary, scaffolds peer support, facilitates higher-order mental processing, and builds knowledge.

In a similar study with older learners, Scott and de la Fuente (2008) investigated pairs of L1-English college students working together to solve Spanish or French grammar problems. Pairs of students were put into two groups. They were given a learning activity that required them to discuss a grammatical rule and co-write an explanation of the rule in the target language. One group was told they could use whatever language they liked to carry out the task (as long as the final product was in

the target language); the other group was told to use only the target language. When the recordings of their conversations were analysed, Scott and de la Fuente found that the interactions of the group who were encouraged to use their L1 were fluid and easy. Pairs multi-tasked, simultaneously reading from the task materials and discussing the text. They discussed the grammar rule, and their discussions were collaborative and balanced. By contrast, discussions in the group in which L1 was prohibited were fragmented, consisted of long silences, and did not address the task effectively. The students "often laughed nervously and looked out of the window" (Scott and de la Fuente, 2008:105). The authors thus concluded that allowing L1 into the discussions facilitated better collaboration, more meta-language, less frustration with the task, and reduced cognitive strain.

A third study (van Weijen et al., 2009) asked Dutch university students to write a short essay in their L2 (English) and to 'think aloud' while they did so. The authors analysed the use of L1. They found that students used their L1 for self-instruction, goal setting, generating ideas, structuring their writing, and meta-commenting (reflecting on the writing process as a whole). These facilitative purposes reflect broadly the findings of both studies discussed above and have been found in other 'think aloud' studies of L1 use in L2 contexts (*e.g.* Wang and Wen 2002, Murphy and Roca de Larios 2010).

The findings of the three illustrative studies here have been observed in broad consistency over a number of other similar studies. All of these find multilingual learners using their L1s to facilitate their engagement in L2 tasks (*e.g.* Storch and Wigglesworth 2003, Alegria de la Colina and del Pilar Garcia Mayo 2009, de la Campa and Nassaji 2009, Kobayashi 2003, Moore 2013, Moodley 2007, Kibler, 2010). Storch and Wigglesworth's (2003:768) conclusion that "the L1 provided essential cognitive support for focusing attention and understanding meaning" is typical of this body of research. Importantly, the authors of these papers tend to stress that this shouldn't mean a linguistic free-for-all in the multilingual classroom, and that teachers should find ways to incorporate L1 into their lessons judiciously, to support learning in the L2.

### 2.8.2 ... But does more fluid co-operation and a lightened cognitive load lead to better task performance?

Illuminating as this body of research is of the functions served by L1, considered on its own it does not provide evidence that allowing or encouraging use of the L1 facilitates improved academic outcomes for those learners. Extrapolating from these observations to suggest that multilingual learners attain more highly in academic tasks when no restriction is placed on the languages they use is a fine theory, but it must be supported by empirical evidence if teachers are to be convinced that changing their policy and practice is worthwhile. Unfortunately, little of the research described above

goes the extra step to determine whether the assumed lightening of the cognitive load and more efficient task management actually resulted in superior academic performance.

That said, two of the above studies did attempt to investigate whether a causal relationship might exist between students' use of L1 and their subsequent academic performance. In Swain and Lapkin's (2000a) study, the authors looked for associations between the quantity of L1 use and the quality of the resulting L2 written pieces. To do this they used a five-point scale to rate the quality of the language and the content in the participants' L2 writing. Based on these ratings, their writing was classified as either 'above median' or 'below median'. Swain and Lapkin then calculated the average proportion of L1 'turns' the pairs took during their discussions. In one task, the average proportion of L1 turns made by pairs producing writing in the 'above median' category was 15%. For pairs producing writing in the 'below median' category the average proportion of L1 turns was 41%. This difference was described as a "significant negative correlation" by Swain and Lapkin (2000a:264). That is, the more L1 the pairs used, the poorer the quality of their L2 writing. In the other task the difference was smaller and non-significant, but nonetheless pairs that took fewer turns in L1 tended to produce better L2 writing.

Swain and Lapkin's findings are reflected in the study by van Weijen et al. (2009). Their analysis was slightly more nuanced than the rather crude method used by Swain and Lapkin. Rather than simply counting the number of L1 turns taken by the participants and assessing this against the binary 'above or below median' classification of the quality of the writing, they classified L1 utterances by the cognitive functions they were assumed to represent. The cognitive functions identified by van Weijen et al. were: self-instruction (*e.g.* "I should start writing something"), goal setting (*e.g.* "I want to summarise what I have just said"), structuring (*e.g.* "Maybe I can turn it into two arguments"), generating ideas (*e.g.* "I'm trying to come up with an introduction"), and meta-commenting (*e.g.* "I'm contradicting myself here") (van Weijen et al. 2009:242). They then calculated the proportion of each type of L1 utterance used by the students. To calculate the quality of the written output, each piece of writing was rated on a 100-point scale. The relationships between type and frequency of L1 utterance and the quality of the students' essays were analysed using structural equation modelling. They found that only meta-commenting was statistically significantly related to L2 text quality, and that this relationship was negative. That is, "writers who make many Metacomments in their L1 while writing in their L2 wrote relatively poor quality L2 texts" (van Weijen et al. 2009:244).

### 2.8.3 Summary of research on how multilingual learners use their L1 in school

Research that explores the functions served by multilingual learners' use of their L1 helps to generate theories of how that language might be incorporated into pedagogy in non-bilingual programmes. For example, this body of research has shown that multilingual learners tend to use their L1 when planning their work. Teachers might, for example, consider explicitly encouraging their multilingual students to plan an essay using their L1 in preparation for writing it in English. This might be an effective way of freeing up cognitive resources, which in turn might allow them to concentrate more fully on the principal learning objectives of a task. Indeed, recent research on this principle (known as cognitive load theory) has been used to make a persuasive case to teachers in all disciplines, not just language teaching, that demands on students' cognitive resources should be focused on the key learning aim of an activity and not be diverted to peripheral cognitive tasks (CESE 2017).

However, despite commentators arguing that explicitly encouraging learners to use their L1 use for these kinds of purpose is time well spent, studies like Swain and Lapkin's (2000a) and van Weijen et al.'s (2009), are insufficient to support that conclusion. Indeed, research in this literature that has attempted to establish whether a relationship exists between quantity of L1 use and quality of the resulting L2 output suggests that overreliance on L1 may not be particularly fruitful.

In order to determine whether explicit use of students' L1 in non-bilingual programmes translates into improved educational outcomes for those students well defined L1-mediated approaches must be evaluated in those terms.

## 2.9 Translating research into the UK context

As we have seen, theories about the way multilingual minds work, and related psycholinguistic research on cross-linguistic transfer, suggests that systematic development of both multilingual learners' L1 and their L2 might be helpful. Research into the cognitive functions performed using L1 by multilingual learners helps us to understand how L1 may scaffold L2 learning and research in bilingual schools shows us what happens to student achievement when L1 and L2 development are both explicitly supported and developed over time.

The final piece of the puzzle is to determine whether the principles that emerge from this body of research can be incorporated into the usual practice of teachers in non-bilingual schools, such as typical mainstream primary schools in the UK. Moreover, if they can be incorporated into usual



practice, does this translate into improved linguistic and academic outcomes for the multilingual pupils in their classes? This benefits from some expansion.

First, a restatement of the purpose of the original research addressed in this thesis. Teachers in linguistically diverse mainstream schools in the UK are advised to make systematic and judicious use of their pupils' L1s, because doing so is thought to confer an advantage in terms of English language proficiency and academic attainment. Some advice to teachers has been somewhat general, for example "Children's experience of different scripts at home should be acknowledged and built on" (QCA 2000:47). Other advice has defined strategies in more detail, for example "Before the Literacy Hour, a bilingual teaching assistant and children talk through the pictures and summarise the story of a Big Book in the home language." (Bourne 2002). Operationalising these suggestions is not without costs to resources of time, effort and money. Teachers need to be assured that such investment is worthwhile in terms of meaningful outcomes for them and their students. For the purposes of this thesis, I have chosen to interpret 'meaningful outcomes' to mean English language proficiency and/or academic attainment. I am, therefore, concerned with assessing the worth of any such investment by teachers, by evaluating the findings of empirical research that addresses advice such as that given by the QCA, Bourne and their likeminded colleagues. I then intend to build on any such evidence through a controlled comparison of an L1-mediated teaching strategy with an alternative strategy that does not use the L1.

### 2.9.1 Limitations of the research

The findings of research in the three broad fields I have described above sets up a relatively compelling basis for following advice to make systematic and judicious use of the L1. For teachers in linguistically diverse mainstream schools, however, the research leaves questions over its validity with regard to their specific context. This is not a fault of the research *per se*, just that none of it appears to have been conducted in linguistically diverse contexts of the sort familiar to teachers in the UK. Additional research is required.

Knowing that relationships exist between similar competences in L1 and L2 suggests that developing one might also contribute to the development of the other. However, as this research is chiefly correlational in nature, in isolation it doesn't reveal evidence of causal relationships. Targeting L1 development might cause improved L2 development, targeting development in L2 might cause improved performance in L1, or a third, 'lurking', variable might be responsible for both L1 and L2 competence<sup>5</sup>.

---

<sup>5</sup> But see Murphy et al. (2014) (Section 2.6.2) for a relatively rare example of an experimental study of this phenomenon that provides

In addition, knowing that multilingual learners refer to their L1 when performing particular cognitive functions, such as planning, structuring and reviewing their work, when they work independently or in collaboration with others who share their L1, suggests that sanctioning such use might be helpful. However, just because someone refers to their L1 when left to their own devices doesn't necessarily mean that formally sanctioning it or explicitly expecting it is the most productive pedagogical approach to take when linguistic and academic development are your goals. As we have seen, such research that exists to address this is not terribly reassuring. Moreover, the point as it relates to the pedagogical decisions made by teachers may be moot. A wealth of research in cognitive psychology on multilingual language processing (e.g. Altarriba and Mathis 1997, Sunderman and Kroll 2006, Blumenfield and Marian 2013, Kroll and Bialystok 2013) tells us that all of the languages known to a multilingual person are always active in their minds. Regardless of what teachers instruct multilingual students to do with their languages, the L1 will always exert its influence over the way they engage in learning tasks.

### 2.9.2 Unpacking the black box of bilingual education

Research on bilingual programmes of education, on the other hand, does provide direct empirical evidence that concurrent use of L1 and L2 is associated with improved L2 proficiency and academic attainment compared with multilingual children who learn only in the target language. The methodologically more robust examples of this research provide a strong basis for the conclusion that attending a bilingual programme is causally related to improved linguistic and academic attainment. But bilingual programmes are not monoliths. They might be thought of as 'black boxes' containing all manner of moving parts. Like all teaching, pedagogy in bilingual programmes is a coalescence of strategies, approaches, demographics, ideologies, environmental influences, and much more. Presumably some of these are responsible for improved linguistic and academic outcomes among students, and some are not. To phrase this in the language I have used to characterise the advice received by teachers in UK schools, bilingual programmes do not so much use the L1 as a pedagogical tool as provide concurrent education using both languages as the media of instruction, with all that that entails. We can be confident in saying that attending a bilingual school is beneficial for multilingual learners, but extrapolating from that to advocate specific strategies for use in non-bilingual schools goes beyond what the evidence tells us. We need to unpack the black box of bilingual education and examine its contents.

---

support for the second of these three potential causal relationships.

Based on conclusions about the effects of bilingual programmes *in totum*, Tikunoff (1985) recognised the need to inquire more deeply to identify what it is about these programmes that makes them advantageous to multilingual learners. This is particularly important if teachers are looking for ways to replicate that advantage in a different educational context. Tikunoff laments the lack of empirical information to describe the “special services and instructional strategies that best meet the linguistic and academic needs for LEP students” and argues that “only occasionally have researchers and practitioners been able to establish concretely what services and instructional strategies work” (Tikunoff 1985:5). To address these shortcomings in the literature, he reports on the *Significant Bilingual Instructional Features (SBIF)* descriptive study carried out under the auspices of the U.S. Department of Education between 1980 and 1982. Fifty-eight elementary school teachers, identified by their peers as exemplary bilingual instructors, were each observed teaching over time. The nature of their interactions with Limited English Proficient (LEP) students during this time were recorded and analysed. The observers identified five instructional features common to all 58 teachers that they considered to be “significant for the instruction of LEP students in terms of obtaining the two goals of bilingual education: (1) acquisition of English proficiency; and (2) [...] acquisition of academic or basic skills proficiency.” (Tikunoff 1985:5).

Perhaps because of the development of our understanding of effective teaching over the intervening three decades, in hindsight three of the effective features identified by the SBIF study seem unremarkable today. These are:

- Active teaching (*e.g.* clear communication of lesson objectives, maintaining student engagement, and monitoring student progress),
- Integration of explicit English language instruction with content instruction, and
- Learning activities that were well matched to the instructional intentions of the teachers.

The other two features identified by the SBIF study were less general, and responded to socio-educational characteristics of the students. These were:

- Use of both the L1 and English for instruction, and
- Sensitivity and responsiveness to the students’ home cultures.

I won’t rehearse the first three features as, while important, they are not germane to the focus of this thesis. The last two features deserve further exploration. In particular, understanding what is meant by ‘use of the L1 for instruction’ is exactly germane to the focus of this thesis and might help shed light on what strategies teachers of EAL learners in UK schools ought to be considering.

Use of L1 by the teachers in the SBIF study was characterised as a tool for mediating understanding of instruction. That is, when children appeared not to understand instruction given in English, their teachers switched to the L1 to clarify. For example, when a child appeared to be struggling to understand what she had been asked to do with a set of counters during a maths lesson, the teacher used Spanish (the child's L1) to work through an exemplar problem with her. After this L1 intervention the child continued to work effectively and independently on the problem she had been struggling with earlier. In another class, when the teacher asked a child who she knew to have very low levels of English proficiency to point to a particular state on a map of the USA, she first gave the instruction in English then repeated it in Spanish. The child carried out the task successfully.

These two vignettes illustrate that for the children in the SBIF study L1 was important, not because it tapped into an otherwise opaque neurocognitive function in which L1 and L2 perform a psychoeducational duet, but because it served the much more prosaic function of ensuring that children with poor English actually knew what they were supposed to be doing. Or, as Tikunoff argues, "These students were able to decode and understand instructional task expectations and new information in terms of the expected instructional outcomes and knew how to achieve them" (1985:7).

The other relevant finding in Tikunoff's study concerned teachers' sensitivity and responsiveness to their students' home cultures. Teachers in the SBIF study "(1) responded to or used L1 cultural referents to enhance instruction, (2) organized instructional activities to build upon ways in which LEP students naturally participate in discourse in their own home cultures, and (3) recognized and honored the values and norms of LEP students' home cultures while teaching those of the majority culture" (Tikunoff 1985:31). Tikunoff points out that the discourse norms of minority groups in a school may differ from those expected by the school. Therefore, teachers who are sensitive to those differences are in a position to understand how classroom activities can be mediated to take them into account. For example, Kale and Luke (1991) describe two conventions associated with oral storytelling in a Torres Strait Islander community living in Australia. One convention is that the person who is 'yarning' (often an elder) does not solicit or tolerate questions while she is doing so. She holds 'speaking rights' until she is done. The other convention is that listeners are not expected to sit attentively while the story is being told, they are free to engage in household tasks or play. This is in marked contrast to the expectations in 'Western' schools, where children must demonstrate that they are actively listening by sitting still, sitting up, and asking and responding to questions and prompts from the teacher as he reads to them. A teacher who is attuned to the expectations placed on Torres Strait Islander children by their home communities is in a position to accommodate those expectations in their classroom practice. Tikunoff (1985) argues, therefore, that framing classroom

instruction and interactions through the lens of students' cultural norms lowers their affective filter (Krashen 1982), makes the tasks feel familiar, and thus facilitates student participation.

Examples of culturally sensitive teaching identified in the SBIF study included teachers of Navajo children ensuring that they did not violate the cultural taboo of girls and boys from the same tribal clan mixing when they assigned children to reading groups; teachers of Hispanic students using Spanish diminutive terms of endearment such as *mijito* (translated by Tikunoff as 'little son') to soften reprimands; and teachers of Chinese children calling on the concept of 'face' (and losing face) to embolden and reassure their students as they were about to take the stage at a school performance for their parents.

Tikunoff concludes this part of the report by saying that "because teachers possessed a second language as well as some knowledge of LEP students' native cultures, they were able to utilize this resource base in order to mediate instruction, and this furthered their students' ability to respond successfully to class task demands" (1985:33).

Of course, there are significant challenges for teachers in linguistically diverse mainstream schools in the UK associated with these findings. The first is that they are unlikely to be proficient in all (or any) of the non-English languages represented in their classes. While some teachers might be in a position to help some children, this is unlikely to be possible for most. Tikunoff's findings are consequently of limited practical value for those teachers. The second challenge relates to how well attuned teachers in UK schools are to the cultural norms of their pupils. For teachers in linguistically (and therefore culturally) diverse schools in the UK, mediating instruction via the cultural norms of their multilingual students does not rely on them being competent in the L1s of those students. Instead it relies on understanding of, and engagement with, the cultures represented in their schools. This feature of successful bilingual education may, therefore, be more straightforward to operationalise than use of the L1 as described above. However, one must surely consider how well attuned it is possible to be to all of the relevant cultural norms represented in such schools before a teacher runs the risk of shallow understanding, stereotyping, or worse "the soft bigotry of low expectations" (Bush 2000).

Tikunoff's (1985) report is a welcome window into the mechanisms of bilingual education. From its findings we have a clearer idea about which elements might be responsible for multilingual students' improved linguistic and academic outcomes. But, as I have argued, the findings are of limited practical value to teachers in the UK.

## 2.10 Stand-alone strategies that use the L1 as a pedagogical tool

Of more practical value for teachers in linguistically diverse schools is research that evaluates clearly described teaching approaches in terms of meaningful linguistic and educational outcomes. Or, as we have already seen Tikunoff (1985:5) describe them, the “special services and instructional strategies that best meet the linguistic and academic needs for LEP students”. Such research should be conducted outside formal bilingual schooling to remove the potential for other important aspects of bilingual programmes to confound the outcomes. These might include, for example, environmental factors such as students attending a school in which the vast majority of their peers share their ethnolinguistic backgrounds, or plentiful opportunities to develop L1 outside of the school, as is the case for English L1 students in French immersion schools in Canada. In addition, it should evaluate these strategies directly in terms of their effects on substantive linguistic and/or academic outcomes. That is to say, research that I consider relevant to addressing this question should go beyond identifying correlational relationships, or reporting on subjective experiences associated with using the L1, to provide an objective measure of the strategies’ impacts in educational terms. While my principal interest is in teaching primary school aged children, I did not restrict literature reviewed in this section to only that conducted with this age group and have included reports of studies conducted with any age group.

### 2.10.1 Observational research

I will begin by describing a case study (Kenner et al. 2008) that explored the use of Bangla and English by third generation Bangladeshi students in a UK primary school during a literacy topic on poetry. I do this to present an example of research that, while on the face of it appears to address the criteria that I have laid out above, is methodologically lacking in some features that I consider to be prerequisites to endowing confidence in the recommendations made by its authors. The prevalence of this kind of research is comparatively high and is regularly used to argue for the educative effects of L1 (*e.g.* Parke et al 2002, Sneddon 2000, García and Sanchez 2015). It therefore deserves some exploration and an explanation of why it does not fully address the question central to this thesis.

Kenner and colleagues’ case study used a traditional Bangla lullaby as a vehicle to explore metaphor with the grandchildren of first-generation Bangladeshi immigrants. The children had family links to Bangladeshi culture, but as a result of having been born in England and having attended mainstream English schooling from the Early Years, their stronger language was English. While most of the children understood the literal meanings of the Bangla lullaby, there were many metaphorical

elements that were opaque to them. This was because understanding these elements relied on specific cultural knowledge that the children did not possess. Through a process of translation from Bangla to English and back again, and discussion of the cultural referents in the lullaby with members of their family, the children began to understand the specific metaphors contained in it.

The authors argue from their observations of the children during the project that extended work around the Bangla poem had positive effects for these children. Their main conclusion was that the activity helped them to connect with their heritage and to strengthen their bicultural identities. This seems an uncontroversial finding, and the sort of claim that can be legitimately made based on the findings of a case study. However, another of the authors' conclusions was to argue that "exploration of stories and poetry through mother tongue as well as English enriches conceptual understanding such as that of metaphor, thus expanding the cognitive dimensions of children's learning" (Kenner et al. 2008:100). In my view this conclusion *is* controversial, or at least, it goes beyond what can be legitimately claimed given the design of the study. If the authors intend to claim that using Bangla alongside English was uniquely responsible for the improvement in understanding of metaphor, in a way that would not be achieved if English was the only language used in the project, they must provide better evidence about the counterfactual position.

Many will argue that the evidence presented by Kenner and colleagues is sufficient to conclude that using the L1 in the way that they did way confers an academic advantage. Indeed, Kenner et al. (2008:100) make just that claim when they say that "the bilingual poetry project demonstrated that the study of literature in school will particularly prosper if all these aspects of children's lives are taken into account" (2008:100). Will it though? With no window on what would have happened had the project been mediated through only English, that claim is difficult to justify. For this reason, I have not included any more studies in which no attempt to provide empirical evidence about the counterfactual was made.

### 2.10.2 Using L1 to prepare for L2 activities

The earliest example of a report on the effects of a stand-alone L1-mediated strategy that I found was conducted thirty years ago among adult Latino students in Boston, USA (Strohmeyer and McGrail 1988). The authors of the report were teachers in a community adult education programme in which one of the explicit aims was to improve English language proficiency among its Spanish L1 students. The project took a learner centred approach to teaching literacy in both Spanish and English using photography as a stimulus for extended writing. Learners were free to choose which language, Spanish or English, they used in their writing. Most chose Spanish. However, the teachers encouraged some of their students to re-write their Spanish essays in English. The authors don't

comment directly on the effect of that strategy on the quality of the English in their report. However, commenting on the project some years later, Auerbach (1993:20) says that as a result of drafting their work in Spanish students went on to "write pieces in English that were considerably more developed than their usual ESL writing".

In my own Master's research I conducted an experimental evaluation of an L1-mediated strategy for helping EAL learners to engage with narrative texts (Chalmers 2014). My research responded to suggestions that EAL learners benefit from talking through the pictures and summarising the story of a book using their L1 with a bilingual teaching assistant before working on it in English later (e.g. Bourne 2002). Thirty-six multilingual learners, representing twelve different L1s, aged between 5 and 7 took part in the study. They were randomly assigned to either L1-mediated or English-only interventions. In the absence of teaching assistants who were proficient in all twelve languages, I asked parents to take on that role. Each child was given a copy of a picture book with a clear narrative structure. The children's parents were given question prompts to help guide them through the process of exploring and summarising the book. Children in the L1-mediated group received these question prompts translated into the relevant L1, while children in the English-only group were given English versions. Parents were asked to use the question prompts to explore and orally summarise the story with their children at home. The children were subsequently asked to write a retelling of the story in English. The quality of their retellings was assessed using the school's usual English literacy assessment framework and a popular commercially available English language assessment tool. The average scores on both of these assessments did not differ statistically significantly between groups. This suggested that the participants had been neither advantaged nor disadvantaged by the language with which they carried out their preview of the book.

### 2.10.3 Systematic comparison of features of L1 with L2

Another controlled comparison of L1 and L2 interventions was conducted with Cypriot upper-primary school children by Yiakoumetti (2006, 2007). The children's L1 was Cypriot Dialect and the language of their formal education was Standard Modern Greek. Yiakoumetti observed that the children often made "interlingual errors" (Yiakoumetti 2007:57) when using Greek. She characterised these errors as those attributable to syntactic, lexical, morphological and phonological features of Cypriot Dialect emerging in the children's Greek discourse. Over three months, children were taught for one session a day using a language comparison approach. This involved systematic comparison of features of Cypriot Dialect with features of Greek to draw participants' attention to differences between the languages that might be responsible for interlingual errors. At the end of the three-month period, participants' Greek was assessed in an oral exam and a written essay, for which the proportion of interlingual errors attributable to Cypriot Dialect was calculated. Their scores were



compared with children in other schools who had not been taught using the language comparison approach, and who had continued as usual with their standard Greek language curriculum. Yiakoumetti found that children taught using the intervention made statistically significantly fewer interlingual errors of all kinds (syntactic, lexical, morphological and phonological) in both writing and speaking than children who had not been taught using the language comparison approach.

In a similar study, Paradowski (2007) conducted a controlled experiment with eight classes of 16 to 17-year-old Polish secondary school children, who were of upper-intermediate English proficiency. Once a week for 45 minutes over nine months, two of these classes were taught English grammar by the researcher using an approach called the Language Interface Method. This method, as with Yiakoumetti's language comparison approach, systematically compared grammatical structures in English with the equivalent grammatical structures in Polish (the participants' L1). The remaining six classes were taught grammar using the preferred methods of their usual teachers "which relied much less on comparisons with the pupils' L1" (Paradowski 2007:160). Paradowski found that students taught using the Language Interface Method statistically significantly outperformed the comparison students in tests of English grammar.

It is interesting that these two studies demonstrate that the principle of comparing languages appears to be effective both in languages that are relatively close (Standard Modern Greek and Cypriot Dialect) and languages that are comparatively distant (Polish and English). This suggests that the principle might be applicable in a variety of different combinations of languages, and thus worth further investigation in the linguistically diverse contexts found in the UK. However, to operationalise this, teachers need to be sufficiently well versed in the grammars of both English and the L1. As we are finding, this challenge is a recurring theme in translating research findings to the UK context.

#### 2.10.4 Using the L1 to give instructions and to teach academic skills

Butzkamm (2003) and Atkinson (1987), in the papers I described in Section 2.2, both suggest that classroom instructions might be more effectively delivered in students' L1 than in their L2. While their motivation was largely for reasons of expediency, research has been conducted to determine whether using the L1 in this way is also beneficial in terms of academic outcomes.

Bozorgian and Pillay (2013) sought to assess whether instruction in L1 might have deeper implications for learning. They investigated how English listening comprehension was affected by teaching adult Iranian EFL students listening skills using their L1. The lessons were conducted in Persian, and students were taught to use techniques such as making inferences, identifying topics, guessing, and note-taking. A comparison group were taught using traditional listening drills using English. The authors found that those taught listening strategies in Persian showed significant

improvement in scores on listening tests compared with those who had been instructed only in English.

Berman (1994) randomly assigned 104 Icelandic 17- and 18-year olds learning English as a foreign language to receive essay writing instruction in either their L1 (Icelandic) or English. The study aimed to assess whether skills newly taught in the L1 transferred to writing competence in English, and if this approach was more effective than instructing them in English only. Responding to the threshold hypothesis, the author was also interested to assess whether “some sort of threshold of language competence ceiling (sic) has to be attained in the L2 before existing abilities in the first language can begin to transfer” (Berman 1994:2). To explore these relationships learners were assessed for their organisational writing skills and grammatical competence at the start of the study. Based on their score they were deemed to be either ‘Less Grammatically Proficient’ or ‘More Grammatically Proficient’. These categories were used subsequently in the analysis. Berman found that the average gain in essay writing scores for all students was very slightly higher for those who had been instructed in their L1. When the analysis took into account participants’ proficiency at the start of the study, for those instructed in their L1 only the ‘More Grammatically Proficient’ students demonstrated greater average gains than participants instructed in English. The average score for participants in the L1 group classified as ‘Less Grammatically Proficient’ was the lowest of all the groups. Berman concluded that skills taught in the L1 do transfer to English writing, but only for students with a sufficiently high level of grammatical proficiency in the target language.

Much contemporary reading instruction relies on explicitly teaching students “cognitive and meta-cognitive skills to help them comprehend text” (Slavin et al. 2009:52). One such approach is known as reciprocal reading. By engaging children in discussion around a text and drawing their attention explicitly to meta-cognitive skills and mental processes such as questioning, summarising, clarifying, and predicting, reciprocal reading aims to help children to “foster and monitor their comprehension” (Fung, Wilkinson and Moore, 2003:1). Using this approach, students work together in small reading groups, with each adopting a specific, defined role in discussions of the reading matter. For example, one student is ‘the summariser’ whose job it is to summarise what has just been read by the group. Another is assigned the role of ‘the predictor’ who predicts what will come next in the text based on the summary given by the summariser. A third is given the job of ‘the questioner’ who raises question stimulated by the text, and so on.

The kind of language needed to discuss meta-cognitive process such as these is complex and abstract. It is the CALP discussed in Section 2.4.2 Students with limited English proficiency but who have developed CALP in their L1 may, therefore, benefit from using their L1 to engage in these

discussions when reading English language texts. In a study of 12 Taiwanese Year 7 and 8 students at three mainstream suburban schools in New Zealand, Fung, Wilkinson and Moore (2003) assessed the impact of reciprocal reading that alternated between Mandarin (the students' L1) and English, on the development of these cognitive and meta-cognitive strategies for reading comprehension in English. The authors interleaved reciprocal reading sessions led in Mandarin with sessions led in English. They noticed that the discussions in Mandarin were far more fluid than those in English, and suggest that this allowed the students to internalise the skills they were being taught more readily, in turn making them available to the reading process in English. They concluded that the "data strongly indicate" (Fung, Wilkinson and Moore 2003:26) that their programme led to improvement in the participants' English reading competence. However, their conclusion must be assessed cautiously as the study did not satisfactorily separate teaching in Mandarin from teaching in English, making it impossible to determine which language was responsible for the improved reading competence they observed.

The findings of these studies support the bottom line conclusion of Tukinoff's (1985) peek into the black box of bilingual education: students who understand what they are supposed to do are more likely to do it better. But the nature of the competences being developed in the studies described above are not prosaic task management skills such as knowing that you are being asked to identify a state on the map of the USA. They likely reflect cross-linguistic transfer of deeper cognitive processes (Cummins 2016).

### 2.10.5 Using the L1 to teach L2 vocabulary

Another area where knowledge and proficiency in the L1 has been successfully used as a learning scaffold is in vocabulary teaching.

Perozzi and Chavez Sanchez (2003) compared the effects of using either Spanish or English to teach English language vocabulary to first grade bilingual children with language delay in their L1. Participants in the study were already attending bilingual schools in the USA, so they were likely to be benefitting already from the advantages associated with that model of education. This study aimed to investigate whether the putative causes of the benefits seen in bilingual education could be leveraged to support children with delayed development in their L1. Thirty-eight first grade Spanish L1 students with delayed Spanish oral language proficiency took part in the study. Their task was to learn a set of 21 prepositions and pronouns in English. The participants were randomly allocated to receive instruction in these words, using a combination of Spanish and English, or using only English. The intervention consisted of teachers showing participants a set of line drawings illustrating the target words, then naming the pronoun or preposition depicted in each. The teacher then said one of the target words and asked the student to point to its picture. When the students had done this

successfully on three separate occasions they were judged to have learned the word. In the Spanish-English group, students were taught the words in Spanish first. When they had correctly identified the target words on three occasions in Spanish, instruction changed to English where the process was repeated. In the English-only group, instruction was delivered in English throughout. To assess the relative effectiveness of each approach, the number of trials (the number of times participants were asked to identify a target word before they achieved three correct responses in English) was compared. For the Spanish-English group this figure consisted of the total number of trials in both Spanish and English conditions. The results showed that students in the Spanish-English group required fewer total trials to demonstrate acquisition of the target words in English than groups taught in English only, by quite some margin.

Another approach to using L1 for vocabulary development involves providing L1 definitions of L2 words during L2 reading activities. A number of studies have evaluated this kind of intervention, all with similar results (e.g. Jacobs, Dufon and Hong 1994; Rott, Williams, and Cameron, 2002; Azari et al. 2012), mostly with adult learners. For example, a study conducted with 135 undergraduate students in Taiwan by Cheng and Good (2009) serves as a good illustration of the general approach and its associated findings.

Cheng and Good (2009) assessed the effects of providing 'glosses' (short definitions) of key words in an English language text on retention of the meanings of those words. Mandarin Chinese speaking students were randomly assigned to one of four different conditions. In one, a glossary was provided in addition to the reading text. It gave Chinese translations of the key words and an example of the word in use in an English sentence. In another condition, Chinese translations of the key words were provided in-line with the text. The key words were printed in bold type, with the translations following them in brackets. In the third intervention the glosses were set in the margin of the text close to where the key word appeared. These consisted of the English word with its Chinese translation. In the final condition, key words were printed in bold type, but no gloss or definition was provided. Students' retention of the words was tested using a gap-fill exercise. They were given English sentences with key words missing and asked to fill the gap using a word chosen from a list of 24 possibilities (the 16 key words and 8 distractors). Their comprehension of the texts was also compared. Students in all three gloss conditions did statistically significantly better on the gap-fill tests than students in the no-gloss condition. Scores on the comprehension test did not differ significantly across groups.

### 2.10.6 Summary of research on stand-alone strategies for using L1

The studies summarised in this section respond to our understanding of the interplay between learners' L1 and L2 to tease out possible stand-alone strategies that could be applied in non-bilingual school contexts. They demonstrate that in some circumstances systematic use of the L1 can be a useful mediator of understanding for multilingual students. Some studies have demonstrated that using the L1 to prepare for work in the L2 might be helpful, and that using it to give procedural instructions and to teach academic skills can be beneficial. We have also seen that using students' L1 as a foil for understanding differences and similarities between their L1 and L2 can improve L2 proficiency. Finally, we have seen that providing L1 translations of vocabulary can help students to understand and learn L2 vocabulary and in some cases to improve comprehension of L2 texts. It is notable that very little of the research that I have been able to identify using non-systematic search strategies has been conducted in primary schools, let alone linguistically diverse primary schools such as those in the UK. Therefore, while the research I have summarised is instructive, if teachers and pupils in these kinds of school are to be confident that adopting L1-mediated strategies will lead to the outcomes that they value, more research is clearly needed.

## 2.11 Summary and conclusion

In this review of the literature I have described an axiom in teaching multilingual learners in mainstream schools that the L1 of multilingual learners can and should be used as a scaffold for their learning in and of a second language. I have outlined the theoretical basis for this axiom, in particular the theories laid out by Cummins in the late 1970s and early 1980s. Cummins' linguistic interdependence hypothesis and Common Underlying Proficiency theory continue to be highly regarded characterisations of the psycholinguistic processes at play when an individual uses more than one language. These theories suggest that proficiencies developed through one language will inform development of the same proficiencies in the other language, so called cross-linguistic transfer. From this it has been argued that maintenance and development of a learner's stronger language while he or she is learning a new language will have a facilitative effect on the learning process. I have described correlational evidence that demonstrates a tendency for positive relationships between L1 and L2 proficiency to exist, and which has been cited in support of Cummins' theories. I have also described research that explores the ways in which multilingual learners use their languages during learning. For example, studies have observed multilingual learners using their L1 for task management, planning and reviewing their work, teaching each other, and for focussing attention. Some have argued that allowing students to use their L1 in this way

lightens the cognitive load, allowing them to focus more fully on the core learning aims of the tasks they are engaged in (Roussel et al. 2017).

While correlational and other observational studies are useful indicators of potential causal mechanisms, taken alone they are not sufficient to make strong claims about the effects of associated pedagogy. To know if deliberate action by teachers to operationalise their pupils' L1 as a pedagogical tool helps to develop academic and linguistic proficiency in L2, experimental research is necessary. The largest body of relevant experimental research has been conducted in bilingual schools. A great deal of this research, over more than four decades, has routinely arrived at the same conclusion: when multilingual children attend bilingual schools they tend to do better (or no worse) on measures of academic and linguistic proficiency than their peers at monolingual schools. In particular, paired bilingual schools, in which pupils are taught in their L1 and L2 in more or less equal measure throughout the school, are associated with the best outcomes for multilingual children.

The evidence gained from bilingual schools provides the necessary empirical evidence to make the claim that concurrent development of L1 proficiency and L2 proficiency is helpful to multilingual learners. However, to extrapolate from this to recommend individual strategies that operationalise pupils' L1 for use in non-bilingual schools (for example using bilingual teaching assistants to interrogate texts with multilingual learners using their L1 before carrying out work based on them in L2) requires its own body of research. Without formal comparisons of these strategies with alternative L2-only strategies to guide them, teachers are not in a position to judge whether the investment of time, money and effort necessary to make the change from their usual practice is a sensible use of these limited resources. There appears to be very little research of this kind in general, and even less of it in primary schools. In particular, there appears to be no research, apart from my own Master's study, that has been conducted with linguistically diverse groups of pupils. That said, of the little research that I have been able to locate, there are some indications that certain strategies might be worth exploring further.

The literature review that constitutes this chapter of my thesis can be considered a narrative review (Gough, Oliver and Thomas 2012), in the sense that I have not specified the methods by which I selected literature for inclusion and that my aim has been to 'tell a story' about L1 use in L2 teaching. The *denouement* of this story is that, while much theorising and related empirical research suggests that primary school teachers might be able to operationalise the L1s of their multilingual pupils to good effect, there is not much in the way of empirical research that would inform how this should be done. It is tempting to conclude that this provides a warrant to embark on new primary

research with the aims of addressing this shortcoming in the literature. However, how can I be confident that my story has included all the relevant literature to guide the direction of that new research, given the way I have selected literature for inclusion? While I have attempted to be fair in my characterisation of the literature, because I have not been systematic in the way that I have selected it, I cannot be certain that I have not missed studies that might provide important information about the state of our knowledge on the question of leveraging L1 as a pedagogical tool in the educational contexts that interest me. If I am to be confident that my primary research will not address a question for which a satisfactory answer is already available, and thus guard against engaging in wasteful research (Chalmers et al. 2014), it is necessary to conduct as systematic and exhaustive a search of the literature as possible. This requires a systematic review (Petticrew and Roberts 2006). A systematic review was the focus of the first phase of the primary research reported in this thesis. It is reported in the next chapter.

# PHASE 1: SYSTEMATIC REVIEW





# CHAPTER 3

## SYSTEMATIC REVIEW

### 3.1 Introduction

In this chapter I will describe what a systematic review is and say why one is needed to inform the direction of my primary intervention research. A systematic review is a piece of primary research in its own right and is reported in accordance with the reporting standards laid out in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement (Moher et al. 2009). These include sections on the objectives of the review, the methods used to prepare the review, the results of the review, and discussion of its findings. In the methods section I describe how I searched for evidence, how I decided whether a study was eligible for inclusion, how I extracted data from individual studies, and how individual studies were assessed for risk of bias. In the results section I provide individual summaries of the included studies and summarise the findings of the body of research as a whole. Systematic reviews may allow meta-analysis of the findings of different individual studies that address similar questions. In this review, studies were not of sufficient similarity, and many did not report data in sufficient detail, to allow for meta-analysis of their results. Instead, I have synthesised the findings using a narrative approach. In the discussion I outline general implications for practice and for future research. I finish by describing how the review informed the direction of my own primary intervention study.

#### 3.1.1 Why a 'systematic' review?

The Achilles heel of any attempt to provide an accurate characterisation of the *state-of-the-art* in sciences research is bias. In traditional approaches to reviewing the literature, reviewers present a compendium of primary studies that have addressed a particular area of enquiry, with the intent of giving the reader an overview of what is known about that area. The problem here is that the reader is not aware of how the authors of such reviews have come to include the evidence in their compendium. When left to our own devices, we reviewers are at liberty to include and exclude studies from the review by any criteria we care to think of. This presents the possibility that reviewers will include only studies of which they are aware, or only studies that conform to their pre-existing prejudices, or only the most recently reported studies, or only studies reported by 'big names', or only studies published in journals with high impact factors, or only studies that were recommended to them by their colleagues, or only studies that ... well, the point needs no labouring.

Left unchecked, these potential biases risk distorting the nature of our collective understanding about an area of enquiry. If a reader takes on faith that a traditional review is an accurate representation of the state-of-the-art, they risk being misled as a result of the conscious and unconscious biases inherent in the manifold unsystematic approaches selecting studies for inclusion.

Being consciously or unconsciously misled by traditional reviews is not merely an academic point. Real world decisions about practice and research are informed by our understanding of the *state-of-the-art*. If these decisions are based on biased information, researchers and policy makers risk, at best, wasting their time and resources pursuing questions for which answers are already known, or missing opportunities to pursue questions of importance where there is uncertainty. At worst, basing practice and research on biased information risks exposing participants in new research to unnecessary risk, or rolling out policy that is useless or actively harmful. One of the chief purposes of systematic reviews, therefore, is to reduce bias in the way that literature is gathered and included so that the conclusions drawn from the body of work as a whole are as trustworthy as possible.

Systematic reviews aim to be as trustworthy as possible by “using systematic and explicit, accountable methods” (Gough, Oliver and Thomas 2012:2). This means that reviewers make explicit the question they are aiming to address, the methods by which evidence is sought, the criteria by which individual studies are included or excluded, and by assessing the individual studies that comprise the review for risk of bias. The most complete reviews also aim to include the totality of relevant evidence. This means that in addition to searching the peer reviewed literature through electronic databases, relevant paper journals are hand searched, and efforts are made to include ‘grey’ literature<sup>6</sup> by searching these repositories, conference proceedings, and by contacting authors in the relevant fields to ask them if they have written, or know of, any unpublished studies that meet the review’s eligibility criteria. While not all systematic reviews are as exhaustive as others in their search strategy, the mark of a systematic review, as opposed to alternative approaches, is that it is open and transparent about the extent to which the literature was searched, and the means by which this was done. This allows the reader to judge from a position of knowledge the completeness of the review in those terms.

---

<sup>6</sup>Grey literature was defined at the Twelfth International Conference on Grey Literature as the “manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by libraries and institutional repositories, but not controlled by commercial publishers; i.e. where publishing is not the primary activity of the producing body.” (Schöpfel 2010).

### 3.1.2 The focus of this systematic review

I conducted this systematic review to generate as complete a picture as I could of what is already known from empirical research about the use of L1 as a 'stand-alone' strategy for teaching multilingual learners in L2 contexts. My narrative review (presented in the previous chapter) identified and summarised theoretical and empirical research that suggests that L1 may be helpful in these contexts. However, information about the methods and effects of doing so in mainstream (non-bilingual) schools, and in particular linguistically diverse mainstream primary schools, was notable by its apparent absence from the readily available literature. This systematic review, therefore, focuses on filling that gap in knowledge by locating, appraising and synthesising research that evaluates teaching approaches that make explicit use of primary and pre-primary school aged multilingual learners' L1 in L2 teaching contexts. Once the extent of our knowledge and the extent of our uncertainty has been established using this method, an informed decision about the focus of my primary intervention study can be made.

## 3.2 Structured Summary

### 3.2.1 Background

Global trends in transnationalism and migration have contributed to a steady rise in the number of children in schools across the world receiving their education in a language other than the one they speak at home. Teachers in such schools have a clear responsibility to adopt strategies that have been shown to be effective in raising attainment for these children. One strategy that has been advanced to meet that end is the 'systematic' use of students' first languages (L1s) as a way to support learning of, and learning in, their second language (L2s). While evidence on the effect of bilingual schools is well known, there appears to be little empirical evidence to describe the effects of adopting approaches that use L1 as a 'stand-alone' teaching strategy in non-bilingual schools.

### 3.2.2 Objectives

The objective of this review was to assess the extent and nature of empirical research into the effects of using primary and pre-primary school bilingual learners' first languages as a pedagogical tool in non-bilingual schools? A secondary aim was to provide an estimate of the direction and magnitude of the effects of adopting L1-mediated teaching approaches in L2 contexts, providing data reported in the studies were of sufficient similarity and adequate trustworthiness to permit this calculation.

### 3.2.3 Methods

*Studies considered for inclusion:* Intervention studies, of any design, conducted in non-bilingual school contexts with children of primary and pre-primary school age, assessing the effects of using students' L1s as a teaching and learning tool, published since 1980, in any language. I used standard methodological procedures expected by The Campbell Collaboration (Campbell Collaboration 2014, 2017) an international organization that prepares, publishes and maintains one of the largest repositories of systematic reviews in the social sciences.

*Information sources:* I searched the following electronic bibliographic databases: British Education Index, Child Development and Adolescent Studies, Education Abstracts, Education Research Complete, Education-Line, Educational Administration Abstracts, ERIC, Academic Research Complete, PsychINFO, Web of Science, SCOPUS, MLA International Bibliography, and SPECTR (1999 version). The most recent search was conducted in April 2015. In addition, I was already aware of a small number of studies that met the review's inclusion criteria.

*Risk of Bias:* Risk of bias was assessed using the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies, which I modified for use with education research.

### 3.2.4 Results

Ten studies met all inclusion criteria and were summarised in the review. The number of participants in these studies ranged from 8 to 443. All studies assessed the effect of using the participants' L1 as a mediating tool for teaching an L2 (in most cases English). The focuses of the studies included vocabulary teaching, reading fluency and comprehension, and general writing skills. The age range of participants in the included studies was 4 to 12 years. Three studies were conducted in the USA, and one each was conducted in Canada, the UK, Italy, Cyprus, The Republic of Korea, Taiwan, and Spain. Based on the EPHPP risk of bias assessment, six studies were judged to be at high risk of bias, two studies were considered to be at low risk of bias, and one study was considered to be at moderate risk of bias. Data of sufficient similarity and detail was not available to synthesise the findings statistically. A narrative synthesis describes contradictory and unclear findings overall, though there is a suggestion that using L1 to enhance instruction of L2 vocabulary might be helpful.

### 3.2.5 Discussion

A paucity of relevant studies, conflicting findings, incomplete reporting of findings, and the small number of studies at low risk of bias means that teachers who are considering using their students' L1s as a pedagogic tool do not have the evidence they need to fully inform that decision process. However, a subset of the studies in this review tend to suggest that using L1 to teach vocabulary in naturalistic settings might be helpful, and should be further evaluated in intervention studies

conducted in a variety of contexts. I conclude that insufficient evidence exists to reliably inform teacher practice in this area.

### 3.3 Aims of the systematic review

In conducting this systematic review I aim to describe the nature and extent of empirical research that compares the effects of L1-mediated teaching approaches with alternative approaches on L2 proficiency and/or academic attainment in primary school-aged children whose L1 is different to the language of their education (their L2).

While teachers are often advised (e.g. Bourne 2002, Celic and Seltzer 2013) that L1-mediated approaches promote success for their bilingual students, both linguistically and academically, the nature and extent of the empirical basis for this advice is unclear.

This review will provide information about:

- the nature of L1-mediated teaching interventions that have been investigated in empirical research
- the contexts in which these approaches have been investigated
- the indicators that are used to compare the effects of alternative teaching and learning approaches
- the study designs used to compare the effects of alternative teaching and learning approaches
- The trustworthiness of each study, principally in terms of efforts to reduce bias
- an estimation of the size and direction of the effects of L1-mediated teaching interventions

#### 3.3.1 Review Questions

The review will address the following research questions:

- What is the nature and extent of empirical research on the effects of using L1-mediated teaching approaches in L2 settings with primary and pre-primary-aged language learners, on outcomes relating to either L2 proficiency or academic attainment, or both?
- What, if anything, can be concluded about the effects of L1-mediated teaching approaches in L2 primary school settings?

## 3.4 Methods

This systematic review was conducted using methods set out in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al. 2009). The PRISMA statement is “an evidence-based minimum set of items for reporting in systematic reviews” (Gough, Oliver and Thomas 2012:80) that helps authors prepare a transparent and reproducible account of their review (Page et al. 2018).

### 3.4.1 Protocol registration

The review title was prospectively registered with the Campbell Collaboration, and is available to view on the Campbell Collaboration website at <http://www.campbellcollaboration.org/library/first-language-mediated-strategies-for-improving-linguistic-proficiency-and-academic-attainment-for-bilingual-children-aged-4-11-in-non-bilingual-schools-a-systematic-review.html>.

### 3.4.2 Information sources

The principal sources of information were bibliographic databases accessed electronically through Oxford Brookes University Library and The Bodleian Library at the University of Oxford. These were:

- British Education Index
- Child Development and Adolescent Studies
- Education Abstracts
- Education Research Complete
- Education-Line
- Educational Administration Abstracts
- ERIC (Educational Research Information Centre)
- Academic Research Complete
- PsychINFO
- Scopus
- Modern Language Association International Bibliography

One additional database that was not available through either of the libraries, or online, was also searched:

- SPECTR (Social, Psychological, Educational and Criminological Trials Register)

After the electronic search had been conducted, the reference sections of eligible studies were scanned and potentially eligible reports contained therein were obtained and screened. This process continued iteratively until no more eligible studies were found. In addition, I was already aware of a small number of potentially eligible studies to which the full inclusion/exclusion criteria were applied.

### 3.4.3 Search

Nomenclature associated with language teaching and learning is extremely varied, often with a number of different terms used to describe similar if not identical concepts. Terminology used in the literature with which I was already familiar provided a starting point to generate a list of primary search terms. These were then entered into individual bibliographic databases' thesauri, if they had one, to identify possible alternatives. The resulting collection of terms was then circulated to colleagues familiar with SLA (second language acquisition) research to solicit suggestions of possible additional terms. The list was also distributed to colleagues experienced in systematic reviewing, and to research librarians at Oxford Brookes University and The University of Oxford to solicit advice about the search strategy more generally. This advice (such as the inclusion of wildcards and the use of limiters) was incorporated into the final search strategy, which is presented in Table 3.1.

In developing a search strategy there is always a tension between sensitivity (sometimes called recall) and precision (sometimes called specificity) (Brunton, Stansfield and Thomas 2012). Sensitive searches aim to identify as high a proportion as possible of potentially relevant material. They cast a wide net by using a large number of search terms, some of which may be 'fringe' terms associated with the research field of interest. Sensitive searches are likely to increase the number of relevant records returned, but in doing so they increase retrieval of irrelevant records, sometimes by orders of magnitude. This has obvious implications for the subsequent task of screening. A precise search, in which searches are limited to a more specific set of terms, results in fewer irrelevant records as a proportion of the whole, but at the risk of finding fewer relevant ones too.

I piloted the search strategy with a sample of databases to assess the nature and quantity of records returned, and to judge whether the balance between sensitivity and precision was appropriate. From this pilot test, it was clear that my search was extremely sensitive, returning a great many irrelevant records. This is not surprising when one considers the ubiquity of terms like 'second language', 'mother-tongue' and 'proficiency' in language research. However, my non-systematic experience of the prevalence of relevant literature suggested that instances of such literature were likely to be vanishingly small. Consequently, I accepted that the more time-consuming task of screening the results was a necessary investment for the increased likelihood of locating what I anticipated to be a



very small number of relevant records. Thus satisfied that the search strategy was returning appropriate records, I applied it to all of the databases.

The search terms were entered into each of three fields in the databases, separated by the Boolean 'AND' function, as one string of terms, separated internally by the Boolean 'OR' function (e.g. L1 OR "first language" OR "mother tongue" etc.) in all databases except SPECTR.

SPECTR is a database of randomised (or possibly randomised) trials of social, psychological, educational and criminological interventions, which was created at the time of the launch of the Campbell Collaboration in 1999. Having attended both the inaugural conferences of the Campbell Collaboration in London and Philadelphia in 1999 and 2000, I knew that the database existed but was unable to find it online. I wrote to members of the Campbell Collaboration to ask if an up to date version was available. The lead author of an article describing the creation of SPECTR (Petrosino et al. 2000) sent me a copy of the original version of the database on CD-ROM. This was the only version that any of the people I contacted were able to locate. SPECTR uses an archaic software package which would not accept multiple search terms. Therefore, I entered each of the terms in Field 1 manually into SPECTR's search bar and tagged the records that were returned. I then repeated the process for the terms in Fields 2 and 3. Finally I cross referenced the tagged records to locate those in which a match for at least one term in each field had been found.

**Table 3.1** Search terms

FIELD 1	AND	FIELD 2	AND	FIELD 3
L1	“	L2	“	proficiency
“first language”	“	“second language*”	“	achievement
“home language”	“	“additional language*”	“	success
“mother tongue”	“	“foreign language*”	“	scores
“community language”	“	“target language*”	“	fluency
“heritage language”	“	“language* of instruction”	“	attainment
“background language”	“	“school language*”		
“ethnic language”	“	“academic language*”		
“native language”	“	“modern foreign language*”		
codeswitching	“	MFL		
“code-switching”				
“code switching”				
“codemeshing”				
“code meshing”				
“code mixing”				
codemixing				
translanguaging				
“language minority”				
“minority language”				
crossing				
polylingual*				
metrolingual*				
multivocal*				
bilanguag*				

Note: each row represents an ‘OR’ function, whereby each term under each field heading (e.g. ‘L1’ OR ‘first language’ OR ‘home language’ OR ‘mother tongue’, and so on) was combined with each term under each of the other field headings using the ‘AND’ function.

The following limiters were applied to the search:

- Fields 1 and 2 limited to abstract only.
- Field 3 not limited (i.e. search terms are included anywhere in the text).
- Whole search limited to publications since 1980, the year in which Cummins elaborated his theory of common underlying proficiency.

The search was conducted in April 2015.

#### 3.4.4 Eligibility criteria

Once the initial search had been conducted, the titles and abstracts were screened for inclusion or exclusion based on the eligibility criteria presented in Table 3.2.

#### 3.4.5 Language of publication

*A priori* exclusion of studies written in languages other than English is not uncommon in some systematic reviews (e.g. Moyles and Yates 2003, Low and Beverton 2004, Nye, Turner and Schwartz 2006). For this review exclusion of studies on the basis solely of their language of publication is not justified. Quite apart from the fact that locating the totality of evidence that meets specific criteria necessarily requires accommodation for publications in languages other than English, research on the language practices of multilingual people is, unsurprisingly, often conducted in areas where the majority language is not English. For example, Canada, South Africa, and Israel all generate a great deal of research on multilingual learners. To assume that this research is only ever published in English (rather than in French, Afrikaans or Hebrew) would be foolhardy.

Of course, non-English publications pose a challenge to monolingual English reviewers. I approached this challenge in the following way. All returned records had an English language version of the title and abstract, regardless of the language of the full publication. If, from screening those titles and abstracts, it looked likely that a study might meet other inclusion criteria, the full report was obtained where possible. Three reports published in languages other than English met this threshold. One report was in French (Nocus et al. 2007), one in Afrikaans (Basson and Le Cordeur 2013) and one in German (Souvignier et al. 2012). I was able to locate and obtain the French and Afrikaans reports, but was unable to do so for the German one.

I translated the full reports of the French and Afrikaans studies using Google Translate. While this is an imperfect translation tool, it gave a reasonable indication of whether each study might be eligible for inclusion. In both cases the Google translation indicated that the study might be helpful, so I sought to obtain a more trustworthy translation. I wrote to the authors of the Afrikaans report to ask if they had written an English language version of it. They had, and they sent me a copy of it (Basson and Le Cordeur 2014). I gave the French language report to a colleague fluent in both French and English, who extracted the relevant data from the report. This allowed me to assess whether the study was eligible or not. In both the Afrikaans and French cases, the studies failed to meet at least one of the inclusion criteria and were excluded.

**Table 3.2** Criteria for inclusion/exclusion in the systematic review

	<b>Include/Exclude Criterion</b>	<b>Rationale</b>	<b>Additional Information</b>
<b>Bibliographic Reference</b>	<b>Include 1:</b> Complete reference	Assessing and locating studies with incomplete or missing titles and abstracts and missing bibliographic information is unlikely to be possible.	Records with incomplete information, such as those with titles only or no bibliographic information, will be excluded. However, if the partial information available indicates that the study might be relevant to the review reasonable efforts will be made to locate the report using other means.
	<b>Exclude 1:</b> Incomplete reference		
<b>Relevance</b>	<b>Include 2:</b> Relevant to assessing the effects of L1-mediated teaching/learning in L2 settings.	Search terms are not exclusive to studies about language learning. For example, L1 and L2 are both terms for lumbar vertebrae. Completely irrelevant reports may be returned as a result.	None
	<b>Exclude 2:</b> Not at all to do with L1-mediated learning in L2 settings.		
<b>Date</b>	<b>Include 3:</b> Published between 1980 and 2015	Cummins' elaborated his Linguistic Interdependence Hypothesis in 1979 and his hypothesis of Common Underlying Proficiency in 1980. These hypotheses are widely regarded as seminal in informing assumptions about L1 and L2 interplay, and the pedagogical implications that follow. Therefore, for the purposes of this review, 1980 is considered the point from which research that responds to these hypotheses can have been conceptualised and conducted.	None
	<b>Exclude 3:</b> Published before 1980		

<b>Design</b>	<b>Include 4:</b> Intervention study	This review seeks to consider empirical, ‘what works’ research on the effects of L1-mediated interventions. Other types of literature will not be included.	This review intends to determine the extent and nature of the empirical basis for advice to teachers to use L1-mediated approaches with language learners. Reports of studies that do not attempt to assess the effects of specific attempts to operationalise theories of linguistic interdependence will not be included.
	<b>Exclude 4:</b> Not an intervention study		A study will be <i>excluded</i> if any of the following apply, and it contains no empirical examples: a methodological study, a study examining relationships and/or statistical associations between variables, theorisation, guidance, a planning document, opinion or exhortation, an editorial, a commentary, a book review, a resource, a textbook, a bibliography, an index or a contents page.
<b>Population and Context</b>	<b>Include 5:</b> Primary/elementary school-aged children.	The catalyst for this scoping review was advice directed specifically at primary school teachers and other educators of primary school-aged children (see Bourne 2002 in section 1.1). The aim of this scoping review is to describe research that can reasonably be assumed to have informed this advice. Research conducted on L1-mediated learning with different demographic groups (such as secondary school-aged children, university students or adult learners) is contextually quite different and will therefore not be included.	Exclude if the mean age of the participants in the study is not within the age range of compulsory primary/elementary education (including preschool) at the time of the study for the country in which the study was conducted. Because the chronological ages of children in primary schools or elementary schools vary slightly from country to country, the classification for the country in which the study was conducted will determine inclusion decision, not the chronological age of the participants.
	<b>Exclude 5:</b> Not primary/elementary school-aged children		
	<b>Include 6:</b> Typical school populations	As for Include/Exclude criterion 5, this review aims to describe research that can reasonably be assumed to have informed advice directed at mainstream school	Exclude if the <i>only</i> participants in the study are children with language disorders, learning difficulties or with special educational needs. Studies conducted in contexts in which

	<p><b>Exclude 6:</b> Studies exclusively of children who have special educational needs, learning difficulties or language disorders.</p>	<p>teachers. Research conducted on L1-mediated learning with children who have special educational needs is contextually quite different and therefore will not be included.</p>	<p>children who have special educational needs may constitute a part of the typical population (e.g. a mainstream school) will not be excluded based on this criterion.</p>
	<p><b>Include 7:</b> Is conducted in a non-bilingual learning context</p> <hr/> <p><b>Exclude 7:</b> Evaluates bilingual schools/programmes or interventions conducted in officially designated bilingual schools/programmes.</p>	<p>The effects of bilingual school programmes have been adequately assessed through a number of systematic reviews and meta analyses (see section 1.2). Additionally, advice such as that presented by Gibbons (2009) and Bourne (2002) (see section 1) is directed specifically at teachers in non-bilingual schools.</p>	<p>Exclude if the study evaluates the effects of bilingual schools or officially designated bilingual programmes of education.</p>
<p><b>Outcomes</b></p>	<p><b>Include 8:</b> Assesses the effects of L1-mediated teaching/learning approaches on L2 proficiency and/or academic attainment more generally as demonstrated through L2, using measures directly related to those outcomes.</p> <hr/> <p><b>Exclude 8:</b> Outcome measure is not of L2 proficiency and/or academic achievement demonstrated through L2.</p>	<p>The use of L1-mediated approaches are recommended because, at least implicitly, it has been suggested that they support linguistic development in L2 and/or that they foster understanding of other curriculum content that ultimately must be demonstrated in L2. Therefore, I will only include studies that report measures of L2 proficiency or that assess understanding of other curriculum content as demonstrated through L2 as outcomes. Surrogate outcomes, such as self-reported satisfaction, perceived levels of engagement, instances of 'off-task' talk, etc. will not be included.</p>	<p>Exclude if a study reports only outcomes that are not direct measures of L2 proficiency and/or academic attainment, or which are surrogates for those outcomes, such as self-reported enjoyment of a task.</p>

## 3.4.6 Study selection

### 3.4.6.1 Phase 1: screening titles and abstracts

Once database searches were complete, returned records were transferred to *Rayyan* (Ouzzani et al. 2016), a web app for screening studies for systematic reviews. Each title and abstract was screened against the inclusion/exclusion criteria. Records that clearly failed to meet one or more of the inclusion criteria were tagged as ‘excluded’, and the reason for exclusion given<sup>7</sup>. Records that could not be excluded based on the information contained in their titles and abstracts were tagged as ‘included’, and their bibliographic information was recorded.

### 3.4.6.2 Phase 2: locating full reports

The bibliographic information from records tagged as ‘included’ in Phase 1 of the screening process was used to locate the full reports. Most of the reports were obtained through online and print journal repositories accessed through the Oxford Brookes and Bodleian libraries. Some journal articles were obtained through authors’ personal web pages or academic websites such as ResearchGate.net. In addition, non-journal articles, such as government reports and items in the newsletters of professional organisations were available at their associated websites. Books were mostly obtained through the Bodleian Library at the University of Oxford and some were available online via Google Books. Some reports required a more creative approach to locate them<sup>8</sup>. The full reports were then screened against the eligibility criteria first by reading the methods section, then, if a clear reason for exclusion had not been identified therein, the text in full.

---

<sup>7</sup> In many cases records could be excluded on more than one criterion. However, usually, only the first exclusion criterion that I encountered was noted.

<sup>8</sup> A report of one study that looked as if it might be eligible for inclusion in the review was published in 1984 in the journal *Acta Paedologica* (Friedenberg 1984). I was unable to locate the article, or indeed the journal, through the two libraries at my disposal, nor through extensive web searching. I decided, therefore, to try to contact the author directly. With only basic bibliographic information available to me, I did not have many clues to help direct my search for the author. I started by searching for her name on Google. A number of Joan Friedenbergs were returned. I scanned these returns for clues that might suggest that the person was either now or previously involved in bilingual education research. Eventually I came across a webpage that described a keynote lecture by a Joan Friedenberg at ‘The First Hector Hammerly Memorial Lecture on Academic Mobbing’, in 2008. I was aware that Hector Hammerly had published on bilingual education in the past. This looked promising. There was no contact information on the website, but there was a picture of Dr Friedenberg. I ran the Google search again, this time in Google Images. I found many photographs of the same person. In addition to the image that had been presented on the Hector Hammerly web page, there were a number photographs of the same woman singing and playing piano in a musical duo. From these photographs, via a raft of local newspaper reports of concerts by a musical act specialising in tributes to the great and good of the American Rock n’ Roll heritage, I was able to find the website of a band called PinkSlip. Here an email address was provided for anyone interested in hiring the band. Still unsure whether the author of the paper that I could not find was the same person as the keynote speaker at the Hector Hammerly lecture, or that that person was one and the same with the singing keyboard player in PinkSlip, I wrote an email, apologising if I had got the wrong person, and enquiring as to whether she was the same Joan Friedenberg who had penned the article in question. Expressing amusement that I had located her through her band’s website, Dr Friedenberg responded that she was indeed one and the same. She provided me with a scan of her article, and was very helpful subsequently, responding to some follow up questions I had about the methodology she had adopted for the study. Ultimately, the study did not meet all of the inclusion criteria and was excluded from the final synthesis.

### 3.4.6.3 Phase 3: citation searching

In the course of reading the full reports of studies returned by the electronic search, I occasionally found citations of studies that looked as if they might meet the inclusion criteria for this review, but which had not been returned by the search. In these cases I obtained the full text where possible and screened the report using the same methods as in Phases 1 and 2. No additional eligible studies were located in this way. In two cases (Ryan 2005 and Wijekumar et al. 2014), authors indicated that additional, potentially eligible studies were about to be conducted, or had been conducted but had not yet been published. I wrote to the lead authors in both cases to ask if these studies had been written up or whether data were available. I received no reply from Dr Ryan, but found the bibliographic information about the follow-up study on his academic resumé online. This allowed me to locate the report. Dr Wijekumar sent me the PowerPoint slides from her presentation at the American Educational Research Association (AERA) annual conference 2016, which described interim results of an ongoing investigation. Neither study met the inclusion criteria for this review.

On completion of Phase 3 the reports that remained formed the body of literature for narrative, and possibly statistical, synthesis.

### 3.4.6.4 Quality assurance

In order to assess the level of reliability of the eligibility criteria, ten percent (21) of the full reports from Phase 2 were read by a doctoral student in the Department of Applied Linguistics at the University of Oxford. She had no knowledge of my decisions about the studies' suitability for the review. Independently, she used the inclusion/exclusion criteria to decide whether each study in the ten percent sample should be included. There was complete agreement between her decisions and mine.

### 3.4.7 Data extraction

Reports of studies that met the eligibility criteria were read in detail and information from them was extracted using a coding sheet created specifically for this task. The coding sheet was used to help summarise the nature of the following aspects of each study.

- 1) Administrative information, such as how the study was identified and its language of publication.
- 2) Contextual information, such as the geographical location of the study, the ages of participants, and the type of school in which it was conducted.
- 3) Descriptive information, such as the research question addressed in the study and the nature of the interventions it compared
- 4) Methodological information about each study, such as the design and size of the study.

The coding sheet, with expansions and examples for each item, is presented in Appendix A.



### 3.4.8 Risk of bias assessment

Risk of bias assessment tools facilitate the examination of reports of investigations for evidence of those studies' trustworthiness. They direct reviewers in the collection and appraisal of information that will give an indication of the quality of the reported study in terms of, for example, the measures investigators have implemented to help reduce bias, the validity and reliability of the assessment tools that have been used, and the extent to which an intervention was faithfully implemented. The more robust the study is in these terms, the more trustworthy one can consider its results.

In many systematic reviews, level of study trustworthiness is used as a criterion for inclusion or exclusion. For example, the What Works Clearinghouse, an organisation in the USA that synthesises and disseminates summary information about studies that "provide credible and reliable evidence of the effectiveness of a given practice, program or policy" (Institute of Education Sciences, no date), has strict methodological inclusion criteria for its reviews. It considers studies without reservation only if allocation to comparison groups has been determined using a "random process" (What Works Clearinghouse 2014:9), and if it has low overall and differential attrition. Such is the nature of research in education that robust, bias-reducing designs such as randomised trials are scarce, particularly in the UK<sup>9</sup>. As a consequence, What Works Clearinghouse reviews can often include only a tiny number of studies (*e.g.* What Works Clearinghouse 2015a) or no studies at all (*e.g.* What Works Clearinghouse 2015b). While these kinds of strict methodological inclusion criteria help to ensure the credibility and reliability of findings on the effectiveness of educational interventions, they can result in a somewhat restricted view of 'the state of the art' in any particular field.

An approach to study selection that acknowledges the desire for robust, bias-reducing study designs, but which also recognises that research using such designs might not always be available is known as a "best-evidence synthesis" (Slavin 1986:5). Slavin describes the underpinning principle of a best-evidence approach, as follows:

"[I]f a literature contains several studies high in external and internal validity, then lower quality studies might largely be excluded from the review [...], or at most briefly mentioned. [...] However, if a set of studies high in external and internal validity does not exist, we might cautiously examine the less well designed studies to see if there is adequate unbiased information to come to any conclusion." (Slavin 1986:6)

---

<sup>9</sup> The first DfE-funded randomised trial in education was published only as recently as 2011 (Torgerson et al. 2011), though elsewhere randomisation has been used as a means to reduce bias in educational intervention studies at least as far back as the 1920s (Walters 1931), and the design has increased in popularity quite rapidly in the past decade (Connolly 2015).

In the absence of studies that can be considered to be of the highest levels of trustworthiness, a best-evidence synthesis may, therefore, allow us to draw some cautious conclusions about an educational intervention or approach. The synthesis might not result in confident conclusions about best practice, but it might help to indicate potentially fruitful directions for future research.

Slavin (1986) suggests that any decision to adopt a best-evidence approach will be informed by expectations about what kinds of research design are likely to be present in the literature of interest. These expectations will be informed by reading reports of existing reviews (systematic or otherwise) and meta-analyses from the field, and by a preliminary review of the literature. As far as I am aware, no systematic reviews of studies that assess the effects of using primary-aged multilingual learners' L1s have been published. Narrative reviews of L1 use in L2 learning, such as Hall and Cook's (2012) *State-of-the-Art Article: Own Language Use in Language Teaching and Learning*, suggest that robust, bias-reducing designs are not common. Turnbull and Dailey-O'Cain's (2009:1) book *First Language Use in Second and Foreign Language Learning*, a volume intended to "unify what is known about the topic", contains only one report of a controlled comparison of alternative teaching interventions (Macaro 2009). In reviewing the literature for another study of L1 use in L2 classrooms (Chalmers 2014), I found no relevant reports that would meet the highest standards of research design for intervention studies. Given this assessment of the published literature, I chose not to use study design as a criterion for inclusion, but instead to use an appropriate tool to help describe and assess the methodological quality of each study that met the inclusion criteria. I used this quality assessment to inform my discussion and to draw conclusions about the strength of evidence of the systematic review as a whole.

### 3.4.9 Choosing an appropriate risk of bias assessment tool

I assessed the quality of the studies included in the review using an adapted version of the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies (J Thomas et al. 2004), hereafter the EPHPP tool. This tool is designed to help assess the quality of studies that collect quantitative data, and can be applied to a variety of research designs.

As the name suggests, the EPHPP tool was designed with quality assessment of public health interventions in mind. The choice to use a tool originally designed for use in the field of public health was taken, in part, because I have been unable to find any suitable, validated tools designed specifically for use with a variety of study designs to assess the quality of education research. Moreover, many guides to preparing systematic reviews in the social sciences specifically recommend quality assessment tools that have been designed for use in fields of healthcare. In their book *Systematic Reviews in the Social Sciences*, Petticrew and Roberts (2006) itemise six tools that

they consider suitable for use with studies that collect quantitative data (the EPHPP tool being one of them). All six of these recommended tools originated in the fields of public health and healthcare. Torgerson (2003), an educational researcher, suggests developing quality assessment procedures based on the Consolidated Standards of Reporting Trials (CONSORT) statement (Moher et al. 2010). CONSORT was designed to improve reporting of randomised trials in healthcare. The Campbell Collaboration, which produces systematic reviews of the effects of social interventions, including education, recommends assessing study quality using the Cochrane<sup>10</sup> Risk of Bias tool (Campbell Collaboration 2014).

While systematic reviews and meta-analyses have been conducted in the field of education for at least as long as they have in medicine (and, in fact, the term meta-analysis was coined by the educational psychologist Gene Glass (1976)), use of these methods to consider the findings of a number of different studies as a whole have become much more commonplace in healthcare than they are in education. Tools to assist this process in health research are commensurately more numerous, and perhaps better developed. Moreover, the EPHPP tool was designed for use with studies of *public health*, as distinct from other fields within healthcare. This distinction is important. Of the various disciplines under the broad umbrella of healthcare from which educationalists might draw useful lessons about the processes involved in systematic reviewing, public health is probably the closest. It strikes me as likely that The UK Faculty of Public Health's definition of the term 'public health' will resonate with educational researchers interested in learning about and promoting effective education; they define it thus: "The *science and art* of promoting and protecting health and well-being, preventing ill-health and prolonging life through the *organised efforts of society*" (Faculty of Public Health 2010, emphasis added). Given the pedigree of systematic reviewing in medicine in general, and the commonalities between public health and education in particular, I believe that my choice to use the EPHPP tool is justified. Nonetheless, it requires some small changes in terminology and some consideration of education-specific issues, a discussion of which follows.

### 3.4.10 Adjusting the EPHPP tool for use with education research

The EPHPP tool asks for information about the following features of a study:

- A) The extent to which the individuals selected to participate were representative of the target population and how many of those selected agreed to participate
- B) The study design and its capacity to control for potential biases

---

<sup>10</sup> Cochrane is "a global independent network of researchers, professionals, patients, carers, and people interested in *health*." (Cochrane 2016).

- C) The measures taken to control for possible confounders (such as random allocation, stratification or matching, or through statistical analysis)
- D) Whether the participants' group allocation was concealed from assessors when they assessed the outcomes and whether participants knew what the research question was
- E) The nature of the tools used to collect data (e.g. standardised psychometric tests or pupil self-report)
- F) The extent of any withdrawal and dropout from the study
- G) Indications of how faithfully the interventions being compared were delivered
- H) Whether analysis was appropriate to the question being asked

For each study, items A to F can be rated as either 'low' risk of bias, 'moderate' risk of bias, or 'high' risk of bias, based on responses to each section. Items G and H are not assigned a rating, but reviewers are invited to comment on the extent to which each principle has been addressed in the study. A global rating based on the ratings given to individual items is then calculated for the report as a whole. A study is considered to have 'low' risk of bias if it has no 'high' ratings, 'moderate' if it has only one 'high' rating, and 'high' if it has two or more 'high' ratings.

Of these eight criteria, two require minor changes to terminology to make them applicable to education research, and two deserve some consideration in view of first principles for reducing bias in intervention studies:

#### 3.4.10.1 Terminology

Item B of the tool uses the term 'controlled clinical trial' to describe comparisons of alternative interventions in which a method other than random allocation has been used to allocate participants prospectively to intervention groups. To make this more appropriate for education research, I have used the term 'non-randomized comparison' instead. Item H asks for the units of allocation and units of analysis, using terms better suited to health research (individual, practice, institution, etc.). I have changed these to more appropriate, comparable terms 'pupil, class/group, school, local authority/school district' (see Appendix B).

#### 3.4.10.2 Sample representativeness

Item A asks for the likelihood that the participants in the study are representative of the target population. This is an extremely difficult, if not impossible, question to answer. Firstly, one needs a clear indication of what the authors of a study consider to be the target population. Then, one needs

to judge how likely it is that participants recruited to the study were representative of that population.

For example, in my study of L1-mediated reading strategies (Chalmers 2014), I invited every one of the EAL learners in Years 1 and 2 at one school to participate. If the target population for that study was deemed to be EAL children in Years 1 and 2 at that school (and had all those invited participated), then I would be justified in describing the sample as representative. However, my research was intended to be useful to teachers beyond that tightly defined context. A more reasonable assumption of what the target population in my study was is EAL learners in England in general. Indeed, that was what I stated the target population in my study to be. The term EAL, however, encompasses such a wide variety of learners as to be almost meaningless without additional qualification. To illustrate this, it is worth reproducing Strand, Malmberg and Hall's (2015) articulation of this issue, from their Education Endowment Foundation report on EAL achievement in the UK:

“The definition of EAL used in the NPD [National Pupil Database] reflects exposure to a language other than English at home or in the community, *it gives no indication of a pupils' proficiency in the English language*. It is important that this is recognised. On the one hand, those recorded as EAL include second or third generation ethnic minority pupils who may be exposed to a language other than English as part of their cultural heritage, but may use English as their everyday language and be quite fluent in it. At the other extreme it includes new migrants arriving in England who may speak little or no English at all, and may have varying levels of literacy in their previous country of origin.” (Strand, Malmberg and Hall 2015:11, original emphasis).

Based on language proficiency alone, therefore, it would require extraordinary confidence to claim to be able to quantify the representativeness of the participants in my study to the population of EAL learners in English schools as a whole. This is before one even considers the other possible differences between these two groups (geographical location, socio-economic status, length of time in the country, and so on).

In the USA, definitions of children who benefit from special provision that takes into account their needs as emerging bilingual learners are somewhat better defined. The term LEP (Limited English Proficient) is applied to children who fall below a threshold proficiency in English, based on formal assessments. This definition is therefore more objective and perhaps more useful than 'EAL', if the intention is to apply findings from one study to the population more generally. However, it still fails

to take account of potential differences that are not based on linguistic proficiency. For example, all of the studies included in this review are conducted in relatively small geographical locations. It is, therefore, unavoidable that any representativeness to LEP children as a whole will be tempered by the effects of the prevailing demographic characteristics in those areas.

For the purposes of this review I chose to assume that authors intended their findings to be applicable to language learners (EAL, LEP or MFL) in general in the country in which the study was conducted, unless they had stated otherwise. Assuming that participants described in the report matched this assumption, and in view of the difficulty of stating categorically how representative any sample is of a target population, I scored Item A (Are the individuals selected to participate in the study likely to be representative of the target population?) as 'somewhat likely'.

#### 3.4.10.3 Control of confounders

Item C of the EPHPP tool asks for an estimate of the extent to which potentially important differences between participants, other than the intervention that they receive, have been controlled. Users of the tool are first asked to indicate whether there were important differences between groups prior to the intervention, then to "indicate the percentage of relevant confounders that were controlled (either in the design (e.g. stratification, matching) or analysis)" (EPHPP tool p2).

Some relevant potential confounders in comparisons might be known to researchers at the time that they design their investigations (age, gender, socio-economic status, and so on), so it may be possible to take account of these through techniques such stratification and statistical matching. However, other potentially relevant confounders will be unrecorded, unknown or unknowable. Item C of the EPHPP tool is, therefore, misleading when it implies that it can be stated categorically whether there were or were not important differences between groups prior to the intervention, and that the percentage of relevant confounders can be estimated. The only way that relevant confounders – known or unknown – can be distributed unbiasedly between comparison groups is by random allocation. Random allocation generates comparison groups between which any differences (known, unknown and unknowable) result from the play of chance and not from systematic differences at the outset. For the purposes of this review, therefore, I have chosen to rephrase Item C, Question 1 as "Were there any known important differences between groups prior to the intervention?". This allows me to acknowledge where authors have made a clear statement of whether they considered their comparison groups to be largely similar or dissimilar on the basis of known characteristics. In addition, I have reduced the answer options in Item C, Question 2, to only Option 1, "80-100% (most)", and Option 4, "Can't Tell". The former was applied only to studies

where comparison groups were generated using random allocation, and the latter for studies that used any other method to generate comparison groups.

With these modifications and considerations taken into account, the EPHPP tool helps to make a reasonable assessment of the trustworthiness of the variety of study designs represented in this review.

### 3.4.11 Incomplete reporting

In a number of cases, authors had not included sufficient information to make a judgment about items on the EPPHP tool. For example, one report did not state whether any steps had been taken to conceal group allocation from assessors, making it impossible to judge whether the trial was at risk of the associated bias. In another report the authors did not describe the method of allocation to comparison groups, making it impossible to judge the extent to which they addressed the possibility of allocation bias. In cases like these, I attempted to contact the authors for clarification. If I did not receive a response the item was scored 'high' on that item. It is quite possible that the authors of these studies had taken appropriate steps to reduce bias on those items, and that I have undersold the quality of these studies. If the authors have neglected to include this kind of information, however, one must err on the side of caution and score the report down. This emphasises the importance of complete reporting of research, and the value of using reporting guidelines such as CONSORT for randomized trials (Moher et al. 2010), STROBE for observational studies (von Elm et al. 2007), and PRISMA for systematic reviews (Moher et al. 2009).

### 3.4.12 Summary measure

Where effect sizes were published in the reports, these were included as summary measures of the effects of (or differences between) the interventions. If effect sizes were not reported but sufficient information was available to allow effect sizes to be calculated, I did this. If sufficient information for these calculations was not reported, then I attempted to contact the authors to request additional data. In the event that authors responded positively to these requests, effect sizes were calculated and reported. If authors did not respond positively, simple differences between intervention groups' scores on the outcome measures used in the study were calculated and reported.

### 3.4.13 Synthesis of results

My intention had been to use the effect sizes published in the individual reports, or to calculate them myself if they were not reported. This would allow statistical synthesis of the results of the included studies, and generate a summary finding for the included literature as a whole. As I will show, insufficient information was available to make these calculations for all studies. Moreover, studies were not of sufficient similarity to make this kind of analysis meaningful. As a consequence, I

adopted a narrative approach to synthesis, whereby the nature and findings of each study is summarised and considered in the light of the nature and findings of the other studies in the review.

#### 3.4.14 Risk of bias across studies

I considered the risk of bias for the set of studies as a whole by assessing the frequency and extent to which individual studies attempted to generate comparable comparison groups, the frequency and extent to which they concealed participants' group allocation from the assessors, and the general appropriateness of the designs to address each study's research question. In the light of this consideration I describe the extent to which the risk of bias has been addressed in the body of work as a whole.

### 3.5 Results

This section documents the process of data collection and quantifies the number of reports located and appraised in each phase, giving reasons for the exclusion of reports as the process progressed. Summaries of each of the studies included in the final review are presented and key features of the body of literature as a whole are summarised.

#### 3.5.1 Study selection

Before conducting the search, I was aware of two studies that met the inclusion criteria (Yiakoumetti 2006, Chalmers 2014). I then conducted the initial search of electronic databases. This returned 5,945 records. Of these, 1,595 were duplicates. Later, by serendipity, I became aware of two other studies that met my inclusion criteria and that had not been returned by the electronic search (Sieh 2008, Camó and Ballester 2015).

Screening the titles and abstracts of the 4,353 electronic records after duplicates were removed yielded 210 records that could not be excluded on the basis of the information reported therein. The 4,143 studies that were excluded at this stage were excluded either because they did not address L1-mediated learning in L2 settings (n=499), they were not intervention studies (n=454), participants were not pre-primary or primary-school-aged (n=922), they were evaluations of or in bilingual programmes (n=972), or they did not assess the effects of the interventions on L2 proficiency or academic achievement as demonstrated through L2 (n=1296)<sup>11</sup>.

---

<sup>11</sup> To protect against the impression that, were it not for differences in demographic characteristics of the participants or a difference in methodological approach this review would have been overwhelmed by studies that might help us understand the effects of using L1-mediated learning approaches in L2 settings, it is important to note that only the first identified exclusion criterion has been recorded. Entirely for reasons of expediency in single-handedly screening close to 6,000 abstracts, once a feature by which a study could be excluded was identified, the study was marked for exclusion, the reason noted, then I moved on to the next record. For example, my search returned a paper entitled *Practitioners' and parents' perceptions and attitudes about bilingual education* (Gill 2013). This could have been excluded on the basis of any one of the following criteria: not an intervention study, not primary/elementary or preschool-aged children,

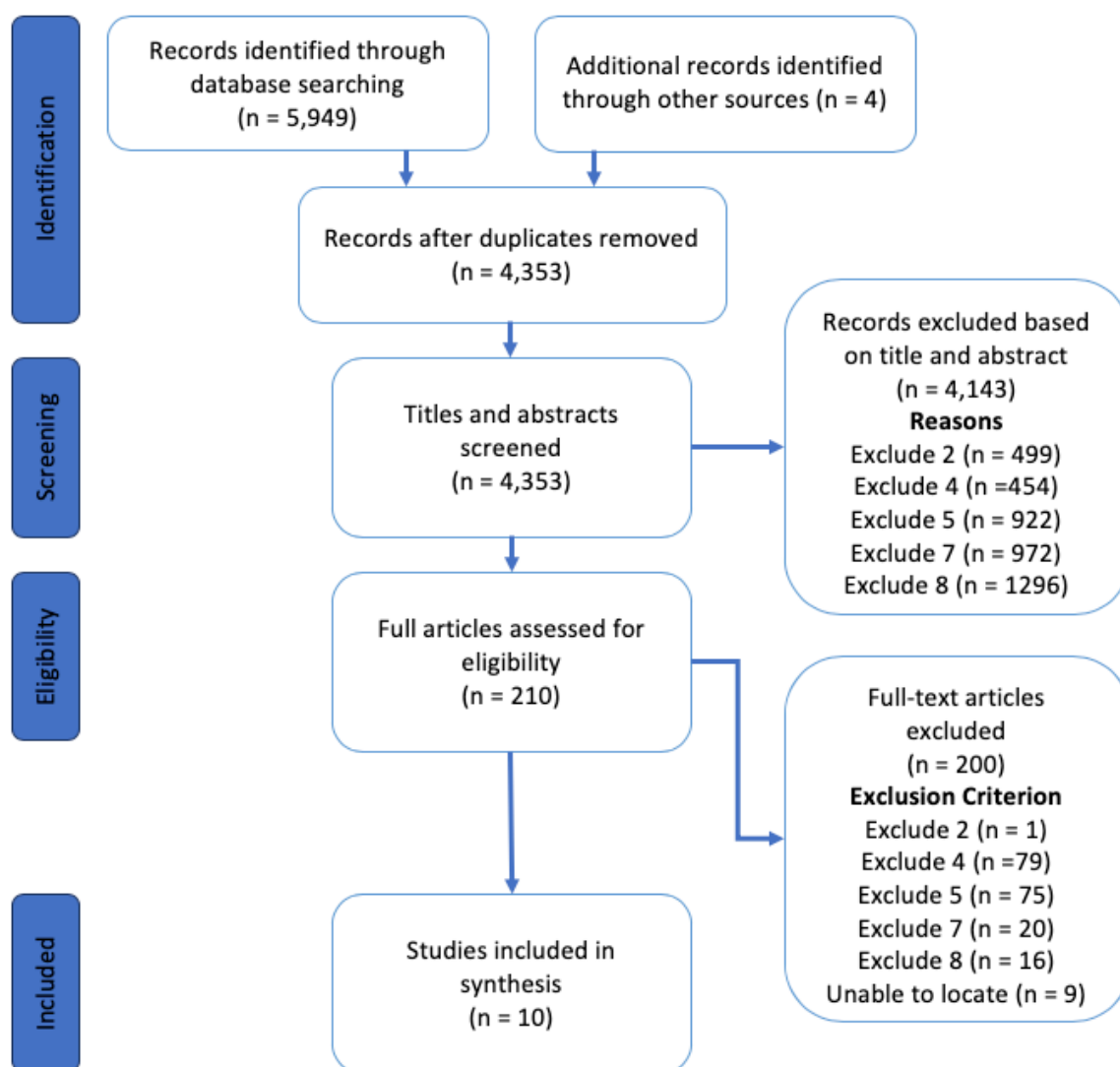


Of the 210 potentially relevant records that remained after phase one screening, I was able to obtain full text versions of 201. Of those, 191 did not meet at least one of the inclusion criteria. The principal reasons for exclusion were: the study did not address L1-mediated learning in L2 settings (n=1), it was not an intervention study (n=79), participants were not primary or pre-primary-school-aged (n=75), it was an evaluation of or in a bilingual programme (n=20), or it did not assess the effects of the interventions on L2 proficiency or academic achievement as demonstrated through L2 (n=16). This left ten studies that met all inclusion criteria and which were included in the final synthesis.

The PRISMA flow diagram (Figure 3.1) describes the screening process and the numbers of records and reports excluded at each phase, and the reasons for exclusions.

---

evaluates bilingual schools/programmes or interventions conducted in bilingual schools/programmes, and outcome measures not of L2 proficiency and/or academic achievement demonstrated through L2. In the event it was tagged for exclusion because it was not of primary/elementary or preschool-aged children, but this tells only a fraction of the paper's story.



**Figure 3.1** PRISMA Flow diagram.

### Key to exclusions

Exclude 2: Not at all to do with L1-mediated learning in L2 settings.

Exclude 4: Not an intervention study

Exclude 5: Not primary/elementary or preschool-aged children

Exclude 7: Evaluates bilingual schools/programmes or interventions conducted in bilingual schools/programmes

Exclude 8: Outcome measures not of L2 proficiency and/or academic achievement demonstrated through L2

### 3.5.2 Included studies

Full references for the included studies:

1. Walters, K. and Gunderson, L. (1985). *Effects of parent volunteers reading first language (L1) books to ESL students*. *The Reading Teacher*, 39:1 66–69.
2. Sánchez, L.Z. (2004). *Effects of Parent Participation Using First Language Curriculum Materials on the English Reading Achievement and Second-Language Acquisition of Hispanic Students*. Unpublished PhD Thesis. Lehigh University, Pennsylvania, USA.
3. Yiakoumetti, A. (2006). *A Bidialectal Programme for the Learning of Standard Modern Greek in Cyprus*. *Applied Linguistics*, 27:2 295-317.
4. Sieh, Y.-C. (2008). *A possible role for the first language in young learners' processing and storage of foreign language vocabulary*. *Annual Review of Education, Communication, and Language Sciences*, Vol.5 136-160.
5. Tonzar, C., Lotto, L. and Job, R. (2009). *L2 Vocabulary Acquisition in Children: Effects of Learning Method and Cognate Status*. *Language Learning*, 59:3 623–646.
6. Lugo-Neris, M.J., Jackson, C.W. and Goldstein, H. (2010). *Facilitating Vocabulary Acquisition of Young English Language Learners*. *Language, Speech, and Hearing Services in Schools*, 41 314–317.
7. Huennekens, M.E. (2013). *The Cross-Linguistic Effects of Dialogic Reading on Young English Language Learners*. Unpublished PhD Thesis. Virginia Commonwealth University, Virginia, USA.
8. Lee, J.H. and Macaro, E. (2013). *Investigating age in the use of L1 or English-only instruction: Vocabulary acquisition by Korean EFL learners*. *The Modern Language Journal*, 97:4 887–901.
9. Chalmers, H. (2014) *Harnessing linguistic diversity in polylingual British curriculum schools. Do L1 mediated home learning tasks improve learning outcomes for bilingual children? A randomised trial*. Masters Dissertation. Oxford Brookes University. Published online by the British Council. Available at [http://englishagenda.britishcouncil.org/sites/default/files/filefield\\_paths/harnessing\\_linguistic\\_diversity\\_for\\_bc\\_v2.pdf](http://englishagenda.britishcouncil.org/sites/default/files/filefield_paths/harnessing_linguistic_diversity_for_bc_v2.pdf) [Accessed 20 June 2016].
10. Camó, A.C. and Ballester, E.P. (2015). *The Effects of Using L1 Translation on Young Learners' Foreign Language Vocabulary Learning*. *Elia*, 15 109–134.

### 3.5.3 Study characteristics

A summary of the basic characteristics of the ten studies included in this review is presented below. This includes an overview of the risk of bias assessment, and structured summaries of each study. Superscript numbers refer to the number assigned to each study (see above).

#### 3.5.3.1 Location

Three studies were conducted in the USA<sup>2,6,7</sup>. Of the remaining seven, one study each was conducted in Canada<sup>1</sup>, Cyprus<sup>3</sup>, Taiwan<sup>4</sup>, Italy<sup>5</sup>, the Republic of Korea<sup>8</sup>, the UK<sup>9</sup>, and Spain<sup>10</sup>. Nine studies either stated explicitly that they were conducted in urban areas, or this can be inferred from descriptions given in the reports. One study described its location as rural<sup>4</sup>.

#### 3.5.3.2 Publication status

Seven studies were published in peer reviewed journals<sup>1,3,4,5,6,8,10</sup>, two were unpublished PhD theses<sup>2,7</sup>, and one was a Master's dissertation that had been published online by The British Council<sup>9</sup>. One study was written up for publication from a Master's dissertation<sup>6</sup> and one from a Bachelor's dissertation<sup>10</sup>.

#### 3.5.3.3 Design

Three studies used random allocation to generate comparison groups or to allocate participants to interventions: two randomised parallel group trials<sup>1,9</sup> and one randomised crossover trial<sup>6</sup>. One study compared the effects of alternative interventions on groups generated prospectively, but not using random allocation<sup>5</sup>. Four studies compared the effects of alternative interventions on two pre-existing groups<sup>3,4,8,10</sup>. One study used a multiple baseline, interrupted time series design<sup>2</sup>. One study measured outcomes of interest a single cohort, before and after the intervention had been implemented<sup>7</sup>.

#### 3.5.3.4 Size

The smallest study involved only eight participants<sup>2</sup>; the largest study involved 443<sup>8</sup>. In three studies the intervention was delivered to individual participants<sup>2,7,9</sup>. In the remaining studies the interventions were delivered simultaneously to groups of children, ranging in size from two or three children to classes of approximately 30.

#### 3.5.3.5 Duration

The duration of the interventions ranged from ten days to three months. In four studies<sup>8,5,3,10</sup> delayed post-tests were administered following the intervention period. The delay ranged from four days to three months.

### 3.5.3.6 Educational programme type

Four studies were conducted in MFL or EFL (Modern Foreign Language/English as a Foreign Language) classes in mainstream primary schools<sup>4,5,8,10</sup>. Two studies were conducted in mainstream primary schools with ESL/ESOL (English as a Second Language/English for Speakers of Other Languages) programmes for children with limited English proficiency<sup>1,2</sup>. One study was in a mainstream primary school with minimal formal support for EAL (English as an Additional Language) pupils<sup>9</sup>. Two studies were conducted in mainstream primary schools where the language of instruction was different to the community language of the pupils<sup>3,7</sup>. One study was in a mainstream pre-school<sup>7</sup>, and one study was conducted with primary school aged participants attending a government funded summer school for migrant workers<sup>6</sup>. Seven studies investigated the effects of the intervention on children of upper primary age range (7 to 12 year-olds)<sup>1,2,3,4,5,8,10</sup>. The remaining three studies investigated pre-school and/or lower primary school-aged children (four to seven year-olds).

### 3.5.3.7 Language/Skills focus

In all but one study<sup>3</sup>, English was the target L2. In one study, German was included as a target language in addition to English<sup>5</sup>. Six studies measured the effects of the interventions on vocabulary skills of some description<sup>1,4,5,6,8,10</sup>. Two studies measured reading comprehension<sup>1,2</sup>, one of which also measured oral reading rate<sup>2</sup>. One study assessed general writing proficiency<sup>9</sup>. One study assessed phonological awareness and alphabet knowledge<sup>7</sup>. The three studies in the USA<sup>2,6,7</sup> focused on children whose L1 was Spanish. The L1s of the participants in the Canadian, Cypriot, Taiwanese, Italian, Korean, and Spanish studies were Cantonese, Greek Cypriot Dialect, Mandarin Chinese, Italian, Korean, and Catalan respectively<sup>1,3,4,5,7,10</sup>. The study in the UK<sup>9</sup> involved children representing 12 different L1s.

### 3.5.3.8 Findings

Of the ten studies, six found in favour of L1-mediated learning over alternatives in which the L1 was not explicitly used<sup>3,4,6,7,8,10</sup>; two did not detect a statistically significant difference between the outcomes of L1-mediated interventions and those of interventions that made no explicit use of L1<sup>1,9</sup>; and two found in favour of interventions that did not make explicit use of participants' L1s<sup>2,5</sup>.

Key information about the studies is summarised in Table 3.3. Statistical data extracted from the included studies are presented in Table 3.4.

**Table 3.3** Summary of key descriptive data extracted from included studies

Reference	Design	Educational Programme type	Skills focus	Number of Participants	Year/grade level, age	Participant Characteristics	Location	Intervention delivered by/to	Duration	Summary result
<b>Walters and Gunderson (1985)</b>	RCT	Mainstream education with ESL programme	Reading: vocabulary and comprehension	39	Grade 4 Ages 9-10	Cantonese L1 in ESL programme	Vancouver, Canada	Parent volunteers/groups of children (n not specified)	Three months	No difference detected
<b>Sanchez (2004)</b>	Multiple baseline interrupted time series	Mainstream education with ESOL programme	Reading: oral reading rate and comprehension	8	Grades 4 and 5 Ages 9-11	Spanish L1, Hispanic, "non-English-proficient"	"Inner-city", USA	Parent/child	Range from 4 to 8 weeks	Favoured No L1
<b>Yiakoumetti (2006)</b>	Non-randomized comparison	Mainstream primary schools	Oral and written production	182	Final Year Primary Ages 11-12	Cypriot Dialect L1	Rural and Urban Schools in Larnaca, Cyprus.	Regular teacher	Three months (with delayed post-test at 3 months later).	Favoured L1
<b>Sieh (2008)</b>	Non-randomized	Mainstream school with EFL programme	Speaking and Listening: Receptive vocabulary	63	Grade 4 Ages 9-10	Mandarin and Taiwanese L1 with average 30	Suburban south Taiwan	Not stated	Two weeks	Favoured L1

	comparis on					months EFL learning experience				
<b>Tonzar et al. (2009)</b>	Non-randomized comparison	MFL	Writing: Expressive vocabulary	123	Grade 4 Ages 9-10	Italian L1. No prior formal MFL learning	Treviso, Italy	Experimenter (regular teacher present)/whole classes (n not specified)	2 weeks (with delayed post-tests 1 week and 1 month later)	Favoured No L1
<b>Lugo-Neris et al. (2010)</b>	Randomized crossover trial	Supplementary summer education programme	Speaking and Listening: Expressive and receptive vocabulary	22	Ages 4-6	Spanish L1 Latino, LEP. Children of migrant workers	“Rural community”, Northern Florida, USA	Researcher/groups of 2-3 children	3-4 weeks	Favoured L1
<b>Heunneken s (2013)</b>	Cohort (one group pre + post)	Mainstream preschool	Reading: Phonological awareness and alphabet knowledge	15	Pre-school Ages 4-5	Latino Spanish L1	“Urban school district”, Virginia, USA	Bilingual research assistant/child	6 weeks	Favoured L1
<b>Lee and Macaro (2013)</b>	Non-randomized comparison	EFL	Reading and Writing: Vocabulary recognition and recall	443	Grade 6 Age 12	Korean L1	Incheon, Korea	Regular MFL teachers/classes of about 30	6 weeks (with delayed posttest)	Favoured L1

									3 weeks later)	
<b>Chalmers (2014)</b>	RCT	Mainstream primary school. Some EAL support	Writing: general	36	Years 1-2 Ages 5-7	Variety of L1s and ethnic backgrounds	Oxford, UK	Parent/child	10 days	No difference detected
<b>Camo &amp; Ballester (2015)</b>	Non-randomised comparison	Mainstream school with EFL programme	Listening: Receptive vocabulary	40	Grade 5 Ages 10-11	Proficient users of Catalan, but some diversity in languages used at home.	Sant Hilari Sacalm, Catalonia, Spain.	Not reported	Two sessions three days apart. (with delayed post-tests 4 days, and one month later)	Favoured L1



**Table 3.4** Summary of the statistical data extracted from the included studies

Reference	Comparator One Mean (SD)*	Comparator Two Mean (SD)*	Comparator Three Mean (SD)*	Difference	Effect Size*
<b>Walters &amp; Gunderson (1985)</b>	<b>Read to in L1</b>	<b>Read to in L2</b>	<b>No intervention control</b>	Diff' in diff' C1:C2, C1:C3	-
	Pre-test mean: 63.23	Pre-test mean: 58.77	Pre-test mean: 58.91		
	Post-test mean: 72.23	Post-test mean: 64.72	Post-test mean: 71.91		
	Difference: 9.00	Difference: 5.95	Difference: 13.00	3.05, 6.05	
<b>Sánchez (2004)</b>	<b>Oral Reading Rate</b>	<b>Oral Reading Rate</b>	<b>Oral Reading Rate</b>	C1:C2, C1:C3	-
	<i>Baseline slope</i> ††	<i>Intervention slope</i> ††	<i>Generalization slope</i> ††		
	Yara 4.00	Yara 0.32	Yara 0.74	-3.68, -3.26	
	David 4.64	David 0.92	David -2.78	-3.72, -7.42	
	Marisol 7.23	Marisol 8.56	Marisol 2.50	1.33, -4.73	
	Javier 7.32	Javier -4.78	Javier -1.28	-12.10, -8.60	
	Elba -0.50	Elba 3.44	Elba 5.04	3.94, 5.54	
	Gloria 6.62	Gloria 9.80	Gloria 0.66	3.18, -5.96	
	Leonel 7.00	Leonel 19.88	Leonel -0.44	12.88, -7.44	
	<b>Comprehension</b>	<b>Comprehension</b>	<b>Comprehension</b>		

	<i>Baseline slope<sup>††</sup></i>	<i>Intervention slope<sup>††</sup></i>	<i>Generalization slope<sup>††</sup></i>	
	Yara 16.50	Yara 0.18	Yara 2.14	-16.32, -14.36
	David 12.68	David 6.60	David 9.96	-6.08, -2.72
	Marisol 1.82	Marisol 13.56	Marisol 8.56	11.74, 6.74
	Javier 14.32	Javier 2.14	Javier -11.42	-12.18, -25.74
	Elba 16.50	Elba -3.82	Elba 5.46	-20.32, -11.04
	Gloria 8.60	Gloria -3.92	Gloria 12.82	-12.52, 4.22
	Leonel 11.61	Leonel -1.14	Leonel 20.00	-12.75, 8.39
<b>Yiakoumetti (2006)**</b>	<b>L1 Programme</b>	<b>Business as Usual Control</b>		-
	<b>Mean oral interlingual error rate</b>	<b>Mean oral interlingual error rate</b>		
	Pre: ≈7 errors per minute	Pre: ≈7 errors per minute		
	Post: <1 errors per minute	Post: ≈7 errors per minute		≈ -6
	<b>Mean written interlingual error rate</b>	<b>Mean written interlingual error rate</b>		
	Pre: ≈7 errors per 100 words	Pre: ≈7 errors per 100 words		≈ -12
	Post: ≈2 errors per 100 words	Post: ≈14 errors per 100 words		
<b>Sieh (2008)</b>	<b>L1 condition</b>	<b>English condition</b>	<b>N/A</b>	-
	<b>Word gain (written test)</b>	<b>Word gain (written test)</b>		

	Story A: 3.55 (3.38)	Story A: 2.03 (1.89)	1.52
	Story B: 5.77 (4.11)	Story B: 5.22 (3.54)	0.55
	<b>Total words correct (online test)</b>	<b>Total words correct (online test)</b>	
	Story A – 19.68 (2.61)	Story A – 19.41 (3.63)	0.27
	Story B – 18.52 (3.16)	Story B – 17.53 (3.63)	0.99
<b>Tonzar et al. (2009)</b>	<b>L2 Word - L1 Word Method</b>	<b>L2 Word - Picture Method</b>	<b>N/A</b>
	<i>English L2</i>	<i>English L2</i>	-
	Immediate Post-test 1: 42%	Immediate Post-test 1: 47%	-5%
	Immediate Post-test 2: 58%	Immediate Post-test 2: 68%	-10%
	Delayed Post-test 1: 42%	Delayed Post-test 1: 53%	-11%
	Delayed Post-test 2: 40%	Delayed Post-test 2: 49%	-9%
	Mean of all post-tests: 46%	Mean of all post-tests: 55%	-9%
	<i>German L2</i>	<i>German L2</i>	
	Immediate Post-test 1: 29%	Immediate Post-test 1: 38%	-9%
	Immediate Post-test 2: 43%	Immediate Post-test 2: 50%	-7%
	Delayed Post-test 1: 27%	Delayed Post-test 1: 36%	-9%
	Delayed Post-test 2: 25%	Delayed Post-test 2: 33%	-8%

	Mean of all post-tests: 31%	Mean of all post-tests: 39%		-8%	
<b>Lugo-Neris et al. (2010)</b>	<b>Vocabulary explained in L1</b>	<b>Vocabulary explained in L2</b>	<b>N/A</b>		
	<i>Mean growth between pre &amp; post test</i>	<i>Mean growth between pre &amp; post test</i>			
	Naming: 1.14 (1.52)	Naming: 0.68 (1.17)		0.46	$\eta^2 = 0.45$
	Expressive definition: 3.41 (3.10)	Expressive definition: 3.0 (2.71)		0.41	$\eta^2 = 0.85$
	Receptive: 3.05 (1.76)	Receptive: 2.73 (2.27)		0.32	$\eta^2 = 0.70$
<b>Huenekens (2013)</b>	<b>Mean score after L1 intervention</b>	<b>Mean score before L1 intervention</b>	<b>N/A</b>	2.46 (3.92)	d = 0.62
	13.47 (5.01)	10.87 (3.11)		$t(14)=2.628,$ $p=0.02$	
<b>Lee and Macaro (2013)</b>	<b>Vocabulary explanations in L1</b>	<b>Vocabulary explanations in L2</b>	<b>N/A</b>		
	<i>Receptive recall</i>	<i>Receptive recall</i>			
	Immediate post-test: 19.06 (15.92)	Immediate post-test: 8.03 (9.79)		11.03	d = 0.835
	Delayed post-test: 7.28 (7.47)	Delayed post-test: 4.93 (6.18)		2.35	d = 0.343
	<i>Receptive recognition</i>	<i>Receptive recognition</i>			
	Immediate post-test: 32.38 (10.67)	Immediate post-test: 24.02 (9.97)		8.36	d = 0.810
	Delayed post-test: 26.15 (8.90)	Delayed post-test: 21.12 (7.92)		5.03	d = 0.597
<b>Chalmers (2014)</b>	<b>Discussion of story in L1</b>	<b>Discussion of story in L2</b>	<b>N/A</b>		

---

<i>English National Curriculum Level Descriptors</i>	<i>English National Curriculum Level Descriptors</i>	
Sentence Structure	Sentence Structure	-
Mean 5.4 (2.3)	Mean 6.4 (1.4)	-1
Median 5.5	Median 6.0	-0.5
Text Organisation	Text Organisation	
Mean 5.3 (2.2)	Mean 6.3 (1.1)	-1
Median 6.0	Median 6.0	0
Composition and Effect	Composition and Effect	
Mean 5.6 (2.5)	Mean 6.3 (1.4)	-0.6
Median 6.0	Median 6.0	0
 <i>WIDA Performance Definitions</i>	 <i>WIDA Performance Definitions</i>	
Linguistic Complexity	Linguistic Complexity	
Mean 3 (1.4)	Mean 3.3 (1.1)	-0.3
Median 3	Median 3	0
Vocabulary Usage	Vocabulary Usage	
Mean 3.1 (1.3)	Mean 3.7 (0.8)	-0.6
Median 3	Median 3	0
Language Control	Language Control	

---

	Mean 3.8 (1.6)	Mean 4.3 (0.9)		
	Median 4.0	Median 4.0	-0.5	
			0	
<b>Camo &amp; Ballester (2015)</b>	<b>L1-mediated vocab learning</b>	<b>English-only vocab learning</b>		
	<i>Immediate Post-test</i>	<i>Immediate Post-test</i>		
	17.33 (1.46)	17.05 (2.78)	0.28	d = .13
	<i>1<sup>st</sup> Delayed Post-test</i>	<i>1<sup>st</sup> Delayed Post-test</i>		
	19.23 (1.34)	15.00 (3.93)	4.23	d = 1.4
	<i>2<sup>nd</sup> Delayed Post-test</i>	<i>2<sup>nd</sup> Delayed Post-test</i>		
	17.48 (2.56)	13.84 (3.89)	3.64	d = 1.1

\* where applicable and stated

\*\* Yiakoumetti does not report these figures in detail, they have been interpreted from the figures (bar charts) used in the report. As such they should be treated with caution.

### 3.5.4 Risk of bias assessment within studies

The risk of bias assessment for each study is incorporated in the summaries that follow this section, with brief commentary. This section gives an overview of the findings of this assessment.

#### 3.5.4.1 Selection bias

Overall, the studies cannot be said to have avoided selection bias in recruiting participants. As I have discussed in my discussion of the EPHPP tool, without a clear definition of what the authors believe their target population to be, it is difficult to assess the extent to which they attempted to minimise selection bias. The strongest study in this respect took a random sample of all ESL children in one school<sup>1</sup>. Of the other studies, the least weak in this respect attempted to recruit a representative sample by attempting to involve *all* children learning the target L2 in a specific age phase in a number of schools or classes<sup>3,4,5,8,9,10</sup>. Of these, only one<sup>9</sup> reports the proportion of those eligible who were actually recruited and who were not lost to follow up. In the remaining three studies participants were selected by the researchers based on several criteria, including test scores and willingness of parents to consent to their children being part of the study. In none of these studies do the authors describe clearly the wider population to which they intend their findings to be applicable. As a consequence of this, all studies have been rated as having ‘moderate’ risk of bias on this criterion.

#### 3.5.4.2 Study design

Three studies stated that they used randomisation either to generate comparison groups<sup>1,9</sup> or to allocate the languages of the interventions and the order in which they were conducted for each participant<sup>6</sup>. Only one<sup>9</sup> described the method by which random allocation was secured, including the measures taken to conceal the schedule from the allocator. Randomisation is the best way to “achieve pre-experimental equation of groups” (Campbell and Stanley 1963:2). Or, more specifically, allocation to conditions based on a well concealed, unbiased allocation schedule generates comparison groups that differ only as a result of the play of chance. As such, these three studies can be considered to have used designs with a ‘low’ risk of bias on this criterion. Five studies<sup>3,4,5,8,10</sup> used a non-random method of allocating participants to alternative interventions, and are also considered by the EPPHP tool to have a ‘low’ risk of bias on this criterion. The remaining studies have been rated moderate on this criterion as they have used less robust designs for supporting causal inference. The least robust study of all<sup>7</sup> used a single group pre/post design in which outcomes were measured in one cohort before and after the intervention, with no comparison group.

### 3.5.4.3 Confounders

It is impossible to say the extent to which confounders have been controlled as, in addition to known possible confounders such as age and gender, there are likely to be unknown confounders, and probably unknowable confounders. In assessing the risk of bias related to control of confounders, I have applied similar principles as for Study Design. Studies that used unbiased allocation schedules<sup>1,6,9</sup> have been rated 'low' risk of bias, and studies that have not used unbiased allocation schedules, or that have not specified how participants were allocated to interventions, are rated 'high' risk of bias.

### 3.5.4.4 Allocation concealment, blinding of outcome assessment and blinding of participants to research questions

Concealing from allocators the schedule used to allocate participants to intervention groups helps to protect against bias in the way that comparison groups are generated. In addition, blinding outcome assessors to participants' group allocation helps to reduce bias in how the assessment is conducted or interpreted. A third related principle is to blind participants from the intervention that they have been allocated. In health research, for example, participants might receive one of two alternative drugs to treat a condition, both of which look identical so that they don't know which of the two they are receiving. In educational research it has been argued (*e.g.* Prideaux 2002) that it is impossible to blind participants to the interventions that they are receiving. The EPPHP tool recognises the desirability of blinding participants to the intervention to which they have been allocated, but also the difficulty in doing this. Instead, as a proxy for this ideal, the tool rates studies as more trustworthy if the specific research questions under investigation are concealed from participants.

Allocation concealment and the method by which participants were allocated to intervention groups was reported in only one study<sup>9</sup>. This study also reported that outcome assessment was conducted blind to group allocation, but that participants were aware of the research question. None of the remaining studies reported the criteria by which participants were allocated to conditions, let alone whether allocation schedules were concealed from allocators. In addition, none of these studies indicated whether assessors were blind to group allocation or whether the research question was concealed from participants. I attempted, therefore, to contact authors on all papers for clarification. Jang-Ho Lee responded to clarify that in his and Ernesto Macaro's study<sup>8</sup> group allocation was concealed from assessors and participants were unaware of the research question. This study is rated as 'low' risk of bias on this criterion as per the EPPHP guidelines. Mirza Lugo-Neris responded to say that in her study<sup>6</sup> children and parents were unaware of the research question, but that assessors were aware of the condition to which participants had been allocated at time of testing.



Her study is, therefore, rated 'moderate'. No other authors responded to this request for clarification. The EPPHP tool states that incomplete reporting about concealment and blinding should be rated 'moderate', therefore this is the rating given to the remaining studies in the review.

#### 3.5.4.5 Data collection

Three studies were rated 'low' risk of bias on the data collection criterion because they used validated, standardised assessment tools<sup>1,7</sup> or described the validation process of researcher-designed tools<sup>6</sup>. One study<sup>9</sup> was given a 'moderate' rating because it used two tools, one that is validated and one that, while officially recognised by England's Department for Education, is widely regarded to be problematic (see, for example, Coe 2014). One study<sup>4</sup> asserted test-retest reliability of the researcher-designed assessment tools, but did not describe their external validity and so has been rated as 'high' risk of bias. The remaining studies used researcher-designed tools, the validity and reliability of which were not discussed in the reports and have, therefore, been rated 'high' risk of bias on this criterion.

#### 3.5.4.6 Withdrawals and dropout

Six studies did not report on dropout<sup>1,3,5,7,8,10</sup> and so were rated 'high' risk of bias on this criterion. There was 25% dropout in two studies<sup>6,9</sup>, so these were rated 'moderate'. One study<sup>2</sup> reported that one of the eight participants dropped out (12.5%), one study<sup>4</sup> reported that one of its 64 participants was withdrawn. Both of these studies have, therefore, been rated 'low' risk of bias on this item.

The risk of bias assessments are summarised in Table 3.5

**Table 3.5** Risk of bias assessments across studies

Reference	Selection Bias	Study Design	Confounders	Blinding	Data Collection Method	Withdrawals and Dropout	Overall
<b>Walters &amp; Gunderson (1985)</b>	Moderate	Low	Low	Moderate	Low	High	<b>Moderate</b>
<b>Sánchez (2004)</b>	Moderate	Moderate	High	Moderate	High	Low	<b>High</b>
<b>Yiakoumetti (2006)</b>	Moderate	Low	High	High	High	High	<b>High</b>
<b>Sieh (2008)</b>	Moderate	Low	High	Moderate	High	Low	<b>High</b>
<b>Tonzar et al. (2009)</b>	Moderate	Low	High	Moderate	High	High	<b>High</b>
<b>Lugo-Neris et al. (2010)</b>	Moderate	Low	Low	Moderate	Low	Moderate	<b>Low</b>
<b>Huennekens (2013)</b>	Moderate	Moderate	High	Moderate	Low	High	<b>High</b>
<b>Lee and Macaro (2013)</b>	Moderate	Moderate	High	Low	Moderate	High	<b>High</b>
<b>Chalmers (2014)</b>	Moderate	Low	Low	Moderate	Moderate	Moderate	<b>Low</b>
<b>Camo &amp; Ballester (2015)</b>	Moderate	Low	High	Moderate	High	High	<b>High</b>

### 3.5.5 Quality assessment across studies

The 'overall' ratings for each study reveal uneven quality for the set of studies as a whole. A majority (seven out of ten)<sup>2,3,4,5,7,8,10</sup> are rated 'high' risk of bias overall, one is rated 'moderate' overall<sup>1</sup>, and two studies are rated 'low' risk of bias overall<sup>6,9</sup>. However, because items on the tool are not weighted, the 'overall' rating taken alone tells us little about what has contributed to each study's quality rating. While each item on the tool is important, I would argue that two items in particular carry most weight when assessing the trustworthiness of a study. These are 1) bias-reducing study designs such as random allocation to comparison conditions, and 2) the extent to which efforts were made to reduce bias by concealing group allocation of the participants from the assessors and blinding participants to the purpose of the study. If these elements of a study's trustworthiness suggest a high risk of bias, the relevance of the quality of other aspects of a study, such as data collection methods, is undermined. That is to say, the quality of data collection method matters little if your study does not compare like with like and your assessors and participants are in a position to (consciously or unconsciously) subvert the study's findings (after Gorard 2014).

When the extent to which bias-reducing study designs have been used is considered in isolation, the body of evidence as a whole is 'moderate' to 'low' risk of bias (seven out of ten are rated 'low'). This suggests that the conclusions of this systematic review are based on generally satisfactory designs for addressing this type of research question, though it is worth noting that only three studies randomly allocated participants to comparison conditions. The extent to which adequate allocation concealment and blinding was carried out across studies is less clear. In most studies information about concealment and blinding was not presented fully, or at all.

### 3.5.6 Individual study summaries

This section presents summaries of each study included in the final review. It gives an overview of each study following the PICO format (participants, intervention, comparator and outcome), and the bottom-line results and authors' conclusions.

## 1. *Effects of parent volunteers reading first language (L1) books to ESL students*

**Reference:** Walters, K. and Gunderson, L. (1985). 'Effects of parent volunteers reading first language (L1) books to ESL students'. *The Reading Teacher*, 39:1, 66–69.

### **Research Question**

Does hearing stories read aloud in Cantonese (the children's L1) interfere with their English reading achievement? (The research question is not explicitly stated in the paper and has been inferred).

### **Participants**

Thirty-nine Cantonese-L1 children aged 9 to 10 were randomly selected to participate from Fourth Grade ESL classes at a large inner-city school with a large proportion of ESL learners in Vancouver, Canada.

### **Interventions and comparisons**

Participants were randomly allocated to one of three intervention groups: Cantonese, English or 'Catch Up'. For forty minutes, twice a week, for three months, parent volunteers read to the Cantonese and English groups from Cantonese- and English-language story books, respectively. The Catch Up group used the same time to engage in "catch up time [where] no formal teaching activities took place" (p68).

### **Outcomes**

Vocabulary knowledge and reading comprehension were assessed before and after the interventions using the 1977 Gates-MacGinitie Reading Test.

### **Results**

No statistically significant differences in the means of the three groups before or after the intervention, as determined by an analysis of variance test, were detected. In addition, a *t*-test comparing scores at the start and end of the study for each group revealed that all participants' scores on the reading test had improved to a degree that was unlikely to be due to the play of chance. That is to say, reading proficiency in each group was similar when they started, improved over the course of the study, and was similar when they finished.

### **Authors' conclusions**

Walters and Gunderson concluded that hearing stories in Cantonese did no harm to the reading proficiency of the participants. They went on to suggest that hearing stories in L1 carried many benefits, such as explicitly valuing the children’s cultural background, promoting wider acceptance of diversity, and demonstrating how non-English-speaking parents can make a positive contribution to their children’s education.

**Comment**

This study was conducted at a time when the prevailing concern about bilingual models of education was whether or not development of the prestige language (in this case English) was negatively affected by the learning or maintenance of another language. In this respect Walters and Gunderson’s study is unique in the context of this review. Their study aimed to find out whether listening to stories read in the students’ L1 could be considered damaging to the development of English, their L2. All other studies in the review were conducted in a time when bilingual education had largely come to be accepted as beneficial (at least in the eyes of Second Language Acquisition research community) and which instead, therefore, address questions of how to maximise learning in multilingual contexts.

Descriptive data, such as the number of participants in each group and the values of standard deviations from the means were not provided in the paper. Attempts to contact the authors for additional data were unsuccessful. Verification of the findings and calculation of effect sizes is, therefore, not possible.

Walters and Gunderson’s suggestion that the Cantonese intervention carried many additional benefits was not addressed in the study, and must be considered speculative.

**Table 3.6** Risk of bias assessment for Study 1

<b>Selection Bias</b>	<b>Study Design</b>	<b>Confounders</b>	<b>Blinding</b>	<b>Data Collection Method</b>	<b>Withdrawals and Dropout</b>
Moderate: participants randomly selected from a pool of eligible participants	Low: random allocation to comparison groups	Low: random allocation	Moderate: not reported	Low: standardized test	High: not reported

## *2. Effects of Parent Participation Using First Language Curriculum Materials on the English Reading Achievement and Second-Language Acquisition of Hispanic Students*

**Reference:** Sánchez L.Z. (2004). *Effects of Parent Participation Using First Language Curriculum Materials on the English Reading Achievement and Second-Language Acquisition of Hispanic Students*. Unpublished PhD Thesis. Lehigh University, Pennsylvania, USA.

### **Research Question**

“Will using the students' curriculum materials, translated into their native language (i.e., Spanish), as an at-home paired reading intervention improve their English reading achievement at differing ESOL levels (i.e., level 1 versus level 2)?” (p18)

### **Participants**

Eight 4<sup>th</sup> and 5<sup>th</sup> grade (ages 9-11) children enrolled in regular classes at an inner-city elementary school in Allentown, Pennsylvania, USA, in a district characterised by the author as approximately 45% Hispanic. Participants were born either in Puerto-Rico or Dominican Republic and had been in the USA for between three months and three years. Selection to participate was based on three characteristics: their ESOL (English for Speakers of Other Languages) teacher's assessment that they were “having difficulties and considered at-risk in reading as they acquired a second language” (p57); a moderate or high level of cognitive academic language proficiency (see Cummins 1980) in Spanish, their L1, as determined by the Woodcock-Muñoz Language Survey; and their parents' capacity to deliver the intervention at home. The eight participants who were selected were split evenly between ESOL level I (new to English) and ESOL Level II (limited English proficiency). One participant dropped out of the study before receiving the intervention, leaving seven participants whose data were analysed.

### **Interventions and comparisons**

The parents of participating children delivered the experimental intervention as follows. Each week a set of reading materials slated for use at school in the following week were translated into Spanish and sent home. Parents were trained in a pedagogic approach known as Paired Reading. Parents were then asked to use this approach to read and discuss the materials with their children at home two to three times over the course of the week, for 15 to 20 minutes at a time. A multiple baseline, interrupted time series design was used to allow a comparison between slope values of outcome

scores in the 'pre-intervention' period with slope values of outcome scores in the 'with-intervention' period.

### **Outcomes**

The principal outcome measure was the participants' oral reading rate in English - the number of words in a passage of text they could read out loud correctly in one minute. Slope values were calculated by comparing the scores from one week to the next to describe the development of each child's oral reading rate. Measurements were taken twice on each occasion, once using the original English-language versions of the translated texts, and once using age appropriate English-language texts which the participants had not seen before, in either Spanish or English. The latter assessment, called a 'generalization' score, was used to determine whether the effects of an intervention that focused on specific texts could be said to transfer to reading proficiency more generally. In addition to oral reading rate, participants were assessed on their comprehension of the texts, the development slopes of which were also calculated.

Finally, average scores for each child on the two outcome measures in the 'pre-intervention' period were compared with average scores in the 'with-intervention' period.

### **Results**

In 19 of the 28 comparisons of 'pre-intervention' and 'with-intervention' slope values, the slope value decreased after the intervention had been implemented.

Mean scores in the 'with-intervention' period were higher than in the 'pre-intervention' period.

### **Authors' conclusions**

Sánchez concludes that the results "do not support the hypotheses that a paired-reading intervention implemented at home in the native language (i.e., Spanish) using classroom materials would significantly improve the students English reading achievement" (p95).

### **Comment**

Sánchez's primary conclusion is sound, based on the data she has presented. However, later in her discussion, she states that "[a]lthough significant results in ORR and comprehension were not observed based on slope metric results after implementation of the paired-reading intervention, good level effects (i.e., mean differences), as reflected by effect sizes (greater than 1.0), were

obtained by most students suggesting that there was substantial growth over time for participants” (p96). This optimistic interpretation of the results has been informed by a comparison of the mean scores for each child in the periods before introducing the intervention and in the period during which the intervention was being conducted. Far from providing any meaningful assessment of the effects of the intervention, however, Sánchez is simply stating that the participants’ reading improved over time. Moreover, given that slope values tended to *decrease* after the introduction of the intervention, that is to say the intervention appears to be associated with a negative effect on the outcomes of interest, it is interesting to speculate on what the character of the overall improvement in reading might have been had the intervention not been implemented at all. In the absence of a control group we are not in a position to say.

**Table 3.7** Risk of bias assessment for Study 2

<b>Selection Bias</b>	<b>Study Design</b>	<b>Confounders</b>	<b>Blinding</b>	<b>Data Collection Method</b>	<b>Withdrawals and Dropout</b>
Moderate:	Moderate:	High:	Moderate:	High:	Low:
Participants selected by the researcher	Within subject differences over time.	No randomisation	Not reported	Validity not described	1 participant (12.5%) dropped out



### *3. A Bidialectal Programme for the Learning of Standard Modern Greek in Cyprus*

**Reference:** Yiakoumetti, A. (2006). 'A Bidialectal Programme for the Learning of Standard Modern Greek in Cyprus'. *Applied Linguistics*. 27:2, 295–317.

#### **Research Question**

Does deliberate comparison of the linguistic forms used in Cypriot Dialect (CD) with those in Standard Modern Greek (SMG) reduce instances of errors in spoken and written SMG production that can be attributed to interlingual errors - errors in SMG that can be attributed to 'interference' from CD forms. (The research question is not explicitly stated in the paper and has been inferred).

#### **Participants**

182 final year primary school students in eight classes at two government primary schools in the Larnaca district of Cyprus participated. Four classes received the intervention and four classes constituted a 'treatment as usual' comparison. Locations of the schools were described as both rural and urban, but it is not clear from the report whether all participants in both schools were allocated either to intervention or comparison groups or whether one school constituted the intervention school and the other the comparison school. Participants all spoke Cypriot Dialect at home and in their communities, and were taught in the formal medium of instruction, Standard Modern Greek.

#### **Interventions and comparisons**

The intervention used a textbook, devised by the author, which formed the basis of a three-month programme of 'Language Awareness'. The programme focused on comparing and contrasting the linguistic forms of Cypriot Dialect and Standard Modern Greek. In particular, differences in the domains of phonology (the sounds of the languages), morphology (the way words are formed), syntax (the arrangement of words and phrases), and lexis (the words of the languages). Participants were given opportunities to read similar texts in both languages and to identify and classify differences between them using the above domains. The programme also involved translating L1 texts into L2, and writing L2 texts in response to visual stimuli. The programme occupied 50% of the time usually allocated to Standard Modern Greek lessons at the school, and was delivered by the participants' usual class teachers. Teachers had been trained to deliver the intervention by the study's author. The comparison group continued with their usual Standard Modern Greek lessons.

## **Outcomes**

The principal outcome was the proportion of errors in the participants' oral and written Standard Modern Greek attributable to interference from Cypriot Dialect. In another, related, paper the author calls these "interlingual errors" (Yiakoumetti 2007:57). Participants were interviewed in Standard Modern Greek by the author about their daily lives. The participants' speech was analysed for the frequency with which elements of Cypriot Dialect phonology, morphology, syntax, and lexis occurred in their discourse. An error rate of Cypriot Dialect occurrences per minute was then calculated for each participant. To assess the effect of the intervention on written proficiency in Standard Modern Greek, the author used essays routinely written by pupils in the participating classes at the end of each school week as part of the schools' formative assessment procedure. These were analysed for the same kinds of error as in the test of oral language. These assessments were administered at baseline, mid-way through the intervention period, at the end of the intervention period, and then again three months after the intervention period had ended.

## **Results**

There were no statistically significant differences in interlingual error rates between intervention and comparison groups at baseline. In the intervention group, a statistically significant reduction in oral interlingual errors occurred between the baseline and mid-programme assessments. There was another statistically significant reduction in interlingual errors between the mid-programme assessment and the immediate post-test at the end of the intervention period. The proportion of interlingual errors did not change statistically significantly between the immediate post-test and the three-month follow-up. This pattern repeated for the written assessments. In the business as usual comparison group, the frequency of interlingual errors in the oral assessment did not change statistically significantly throughout the intervention and follow-up. In the written assessment there was a statistically significant increase in interlingual errors between the baseline and mid-programme assessment. Subsequently there was no statistically significant change in error rate from the mid-programme, to immediate, and then delayed post-tests. No precise data are provided to illustrate these differences. The author states that the differences were or were not statistically significant and provides the p-value.

## **Authors' conclusions**

Yiakoumetti concludes that "the project confirmed that the ability to consciously identify differences between two [linguistic] varieties enhances performance in the variety which is targeted for improvement" (p312).

## Comment

The study uses that term ‘dialect’ to characterize the two linguistic varieties used by the participants in this study. That is, Cypriot Dialect and Standard Modern Greek were characterised essentially as two versions of the same language, and thus may not qualify as L1 and L2 as I have considered the terms in setting the scope for this review. I must note therefore that, while some people differentiate between dialects and languages, for the purpose of this review I have chosen to align myself with those who do not see the distinction (*e.g.* Otheguy, García, and Reid 2015). The difference between two languages, or between two dialects, or between a language and a dialect is a political construct, rather than a linguistic one. Or as sociolinguist Max Weinreich put it “A language is a dialect with an army and a navy” (1945:13).

**Table 3.8** Risk of bias assessment for Study 3

<b>Selection Bias</b>	<b>Study Design</b>	<b>Confounders</b>	<b>Blinding</b>	<b>Data Collection Method</b>	<b>Withdrawals and Dropout</b>
Moderate:	Low:	High:	High:	High:	High:
Participants based on their teachers, who were well disposed to introducing Cypriot Dialect into their classrooms.	Non-randomised comparison	No randomisation	Group allocation not concealed from assessor. Assessor was the author of the intervention programme text book.	Validity not described	Not reported.

#### 4. *A possible role for the first language in young learners' processing and storage of foreign language vocabulary*

**Reference:** Sieh, Y.-C. (2008). 'A possible role for the first language in young learners' processing and storage of foreign language vocabulary'. *Annual Review of Education, Communication, and Language Sciences*. 5, 136-160.

##### **Research Question**

What are the differential effects of using L1 translations to teach English language vocabulary items compared with using picture cues? (The research question is not explicitly stated in the paper and has been inferred).

##### **Participants**

64 Fourth Grade elementary school children in a school in a suburban location in southern Taiwan. Mean age was 9 years 5 months. Language of instruction in the school was Mandarin Chinese. The author reports that "Up to 60% of them used a mixture of both Taiwanese and Mandarin Chinese at home" (p147), but does not state what language was used by the remaining 40%. Participants had a mean of 30 months' experience learning English as a Foreign Language. None had visited an English speaking country. One participant was withdrawn from the study prior to outcome assessment.

##### **Interventions and comparisons**

Intact classes of children were allocated to one of two vocabulary learning conditions, either word-learning with L1 translation, or word learning using only pictures, and no L1. The method of allocation is not described in the report, and attempts to contact the author for clarification were unsuccessful. In both conditions the intervention was delivered *en masse*. The classes were read to from two story books over four sessions (two sessions for each book). The teacher drew attention to the target items of vocabulary in each book as they were encountered during reading and explained their meanings using flash cards with the written word and a picture representing it. In the L1 condition the children were also told the Mandarin Chinese translation of the word. Otherwise all instruction was in English.

##### **Outcomes**

Receptive understanding of the target items was assessed. Two separate outcome assessments were conducted for each book. One was a paper and pencil test in which pictures of each item were presented to the participants. An audio recording of the target words being spoken was played, with each word being given a number. The participants wrote the numbers of the spoken words next to

the pictures they felt best matched them. This was done as a pretest and post-test, and the average difference in scores on these tests for each group was reported as 'word gain'. The other test was conducted online. The test consisted of 26 assessed words for each book. Participants were shown a picture on screen and were played an audio recording of a word. They report states only that "subjects were allowed six seconds to right-click or left-click the mouse to indicate their answer to each question" (p150), but does not give any more information about what the right and left click indicated. Attempts to contact the author for clarification were unsuccessful. The online tests were administered only once, post intervention. Thus no 'word gain' has been calculated on this measure, only post-test averages for each group.

## **Results**

For both stories, in the pencil and paper version of the outcome assessment, the class in the L1 condition made greater 'word gains' than the class in the English-only condition. Again for both stories, on the online assessment the mean score of the class in the L1 condition was higher than the mean score of the class in the English-only condition. Only in the paper and pencil test for Story A were these differences statistically significant.

## **Authors' conclusions**

The author states that "Provided with L1 glosses in the explicit vocabulary teaching stage, the control group was able to gain more new words and close the statistically significant gap in vocabulary knowledge for Story A" (p153), suggesting that using the L1 to translate the words is more effective than using only English. The implications to why this was not the case for Story B, and for the online tests for both stories, are not explored.

## **Comment**

While the statistically significantly greater gains for L1-mediated learning in one of the four outcomes assessed provides support for the hypothesis that using L1 is helpful, the statistically non-significant differences in the other three outcomes measured are difficult to reconcile with this. Taken as a whole this might be interpreted to mean that, in the main (three out of four times), both L1-mediated and English-only approaches are apparently equally effective.

**Table 3.9** Risk of bias assessment for Study 4

<b>Selection Bias</b>	<b>Study Design</b>	<b>Confounders</b>	<b>Blinding</b>	<b>Data Collection Method</b>	<b>Withdrawals and Dropout</b>
Moderate: Whole classes in one school selected.	Low: Non-randomized comparison	High: No randomisation	Moderate: Not reported	High: Researcher developed tool, validation not described	Low: One child of the initial 64 withdrew from the study.

## 5. L2 Vocabulary Acquisition in Children: Effects of Learning Method and Cognate Status

**Reference:** Tonzar, C., Lotto, L. and Job, R. (2009). 'L2 Vocabulary Acquisition in Children: Effects of Learning Method and Cognate Status'. *Language Learning*, 59:3, 623–646.

**Research Question:** Do primary-school-aged students learn new foreign language vocabulary items more effectively if they are presented concurrently with L1 translations or picture representations of them, and is this affected by the cognate status of the words being learned? (The research question is not explicitly stated in the paper and has been inferred).

### Participants

One-hundred and twenty-three Grade 4 (ages 9-10) school children in Treviso, Italy, whose L1 was Italian and who had not previously engaged in any formal Modern Foreign Languages (MFL) learning.

### Interventions and comparisons

Intact classes of children were allocated to one of two vocabulary learning conditions, either a 'word-learning method' or a 'picture-learning method'. The method of allocation is not described in the report, and attempts to contact the authors for clarification were unsuccessful. In both conditions the intervention was delivered *en masse*. Using a projector, the class was shown an English or German word presented alongside either an Italian translation of that word or a pictorial representation of it, for eight seconds. The experimenter read the foreign word aloud to the participants. A total of forty word-word or word-picture pairs, twenty from each language, were presented in a randomised order. This was repeated twice more, each time with the words presented in a new randomised order. One week after the first session, the entire process was repeated.

### Outcomes

Participants were tested on their recall of each word. This was done by presenting them with either the Italian word or the pictorial representation used in their learning sessions for twelve seconds, then asking them to write down what they thought the corresponding foreign word was. This was conducted immediately after each session, then again one week after the second session, then once more, one month after the second session.

### Results

In all cases, the percentage of words correctly recalled was statistically significantly higher for participants in the word-picture condition.

In addition, cognate words (words with the same meaning that sound or look similar, such as the Italian *triangolo* and the English ‘triangle’) were better recalled than non-cognate words (words with the same meaning, that do not look or sound similar, such as the Italian *pozzo* and the English ‘well’). Also, English words were better recalled, on average, than German words.

### Authors’ conclusions

Tonzar et al. conclude that “[i]n general, the picture-learning method led to better performance than the word-learning method. The superiority of the picture-learning method holds for both L2s” (p 635).

### Comment

The authors report effect sizes, expressed as  $\eta p^2$  (partial eta squared), to describe the differences in effect of the word-picture condition and the word-word condition. The differential effect size of the word-picture condition for English words was 0.08 and for German words was 0.10. Both of these are considered ‘moderate’ (Draper 2018). Data that would allow for verification of these findings (standard deviations from the mean of each group, numbers of cases in each condition, and the unit of allocation to intervention groups) are not presented. Attempts to contact the authors for clarification were unsuccessful.

**Table 3.10** Risk of bias assessment for Study 5

Selection Bias	Study Design	Confounders	Blinding	Data Collection Method	Withdrawals and Dropout
Moderate:	Low:	High:	Moderate:	High:	High:
Whole classes and a range of schools recruited, representativeness dependent on how target population is defined	Non-randomized comparison	No randomisation	Not reported	Researcher developed tool, validation not described	Not reported



## 6. Facilitating vocabulary acquisition of young English language learners

**Reference:** Lugo-Neris, M.J., Jackson, C.W. and Goldstein, H. (2010). 'Facilitating vocabulary acquisition of young English language learners'. *Language, Speech, and Hearing Services in Schools*, 41, 314–317.

**Research Question:** “Do Spanish-speaking children with LEP [limited English proficiency] benefit more from an English-supplemented-with-Spanish direct vocabulary instruction than an English-only instruction during English shared book reading?” (p316).

### **Participants**

Twenty-nine children, aged between four and six years, attending a summer education programme for the children of migrant workers in Northern Florida, USA. All participants were considered Spanish dominant and were classified as Limited English Proficient (LEP) based on scores on a standardised test of single word receptive vocabulary. Seven children dropped out before completion, leaving 22 participants whose data was included in the final analysis, eleven boys and eleven girls.

### **Interventions and comparisons**

In a randomized crossover trial, the effects of two intervention conditions were compared, one in which the meanings of preselected vocabulary items were explained to participants in Spanish and one in which they were explained in English. In both conditions, groups of two or three children at a time were read to by bilingual research assistants from one of four English language storybooks. As each target vocabulary item was encountered, the research assistant explained its meaning using either Spanish or English. The language in which each child heard the explanations for each book was determined by random allocation. In addition, the order in which each child heard the books was randomly determined. Each session took between 15 and 20 minutes and was repeated three times over the course of a week. This entire process was repeated for four weeks, until all participants had been read to from all four books.

### **Outcomes**

Participants' scores on a researcher-designed probe of expressive and receptive knowledge of the target vocabulary items were used to assess the effects of the interventions. The probe assessed three elements of vocabulary knowledge: English naming (in which participants were asked to label a picture in English); expressive definition (in which participants were asked to define the vocabulary

item, using either English or Spanish); and receptive vocabulary (in which participants were shown three pictures and asked to select the one that matched a word being said by the assessor). The assessment was conducted once at baseline, using all of the target vocabulary items, then again at the end of each week, using only the items specific to that week’s book.

## Results

The authors report a statistically significant interaction between language of instruction and growth in scores on the expressive definition portion of the probe. Specifically, when the meanings of the English words were explained to participants in Spanish, they were subsequently better able to define them, in either language, compared with words that had been explained to them in English. There were no statistically significant interactions between language of instruction and measures of English naming or receptive vocabulary.

## Authors’ conclusions

Lugo-Neris and colleagues conclude by suggesting that “[v]ocabulary bridging to the child’s strongest language is a promising strategy to teach novel words to dual language learners” (2010:323).

**Table 3.11** Risk of bias assessment for Study 6

<b>Selection Bias</b>	<b>Study Design</b>	<b>Confounders</b>	<b>Blinding</b>	<b>Data Collection Method</b>	<b>Withdrawals and Dropout</b>
Moderate:	Low:	Low:	Moderate:	Low:	Moderate:
Participants were selected, parental consent listed as first eligibility criterion.	Randomized crossover trial	Random allocation to languages for each ‘phase’.	Participants and their parents unaware of the research question. Assessor not blinded to group allocation.	Scores on researcher-developed tool and a validated standardised tool showed strong positive correlation	25% did not complete the study

## *7. The Cross-Linguistic Effects of Dialogic Reading on Young English Language Learners*

**Reference:** Huennekens, M.E. (2013). *The Cross-Linguistic Effects of Dialogic Reading on Young English Language Learners*. Unpublished PhD Thesis. Virginia Commonwealth University, Virginia, USA.

### **Research Question**

“Does Dialogic Reading in young DLLs’ [Dual Language Learners] home language (L1) improve their phonological awareness skills and alphabet knowledge in L2 (English)?” (p11).

### **Participants**

Fifteen children aged four to five years who were Spanish/English dual language learners were recruited from six preschool classes in three schools in an ethnically diverse urban school district of Virginia, USA. Participants’ families had migrated to the USA from a variety of Central and South American countries, some before their participating child had been born and some after. On recruitment, all participants fell into either ‘below average’ or ‘average’ range on standardised tests of Spanish and English language skills.

### **Interventions and comparisons**

The intervention consisted of five 20-minute sessions per week of Dialogic Reading (a strategy in which children read with an adult who uses prompts and questions to scaffold and extend the child’s engagement with a text). The study lasted six weeks, and was conducted in the school by bilingual Spanish/English research assistants. During these sessions the research assistant read to one child at a time from Spanish language storybooks, stopping to highlight, discuss and provide instruction relating to phonological awareness and alphabet knowledge of target letters contained in the text. All interactions between the child and the research assistant were conducted in Spanish. The participants’ performance at the beginning and end of the study were compared.

### **Outcomes**

The participants’ phonological awareness, described by the author as the “ability to detect, manipulate, or analyse the auditory aspects of spoken language (including the ability to distinguish or segment words, syllables, or phonemes), independent of meaning” (Huennekens 2013:12), was assessed using the *Get Ready To Read! Screening Tool – revised*.

## Results

A paired-samples *t*-test was used to compare scores before and after the intervention. It found a statistically significant increase in phonological awareness between assessments.

## Authors' conclusions

The author asserts that the Dialogic Reading “may have increased the participants’ English language emergent literacy skills” (Huennekens 2013:78).

**Table 3.12** Risk of bias assessment for Study 7

<b>Selection Bias</b>	<b>Study Design</b>	<b>Confounders</b>	<b>Blinding</b>	<b>Data Collection Method</b>	<b>Withdrawals and Dropout</b>
Moderate:	Moderate:	High:	Moderate:	Low:	High:
Sample was not random, selection protocol not stated	Single group pre and post	No controlling for confounders attempted	Not reported	Standardized test	Not reported

## Comment

The primary aim of the study reported in this doctoral thesis was to assess the effects of Dialogic Reading conducted in Spanish, children’s L1s, on their phonological awareness in that language. The portion of the study summarised here relates only to a secondary aim of the study, which was to assess whether there was also an associated effect on the children’s phonological awareness in English, their L2. For reasons that are not made clear in the report, the method for addressing the secondary aim was much less rigorous than those used to address the primary aim. The development of phonological awareness in Spanish was assessed using a multiple baseline, interrupted time series design, in which participants’ awareness of Spanish phonology was measured at regular intervals before and after the introduction of the intervention. This allowed slope values to be calculated to describe the progress made by each child over time. The expectation in this kind of design is that different relationships between the interventions and the outcome measure will be reflected in changes to the values of the slopes following the introduction of the novel intervention. However, despite having set up the experiment in such a way that, at least theoretically, Huennekens could have collected data about English phonological awareness in the same way and with the same diligence that she did for Spanish phonological awareness, she chose not to. As a result we have only average scores on pre- and post-test measures of English phonological awareness in one group of children, each of who received the intervention for a different length of time (because of the

multiple baseline element of the design). Even allowing for the hedging language of Huennekens' conclusion, it is difficult to confidently accept any assertion that Dialogic Reading alone (or indeed, at all) was responsible for the change observed.

## 8. Investigating age in the use of L1 or English-only instruction: Vocabulary acquisition by Korean EFL learners.

**Reference:** Lee, J.H. and Macaro, E. (2013). 'Investigating age in the use of L1 or English-only instruction: Vocabulary acquisition by Korean EFL learners'. *The Modern Language Journal*, 97:4, 887–901.

### Research Question

“What are the relative effects on English vocabulary acquisition and retention of young EFL [English as a Foreign Language] learners resulting either from teacher code-switching to Korean or from giving English-only explanations?” (p890)

### Participants

443 Grade 6 (12 year old) pupils studying English as a Foreign Language (EFL) in two state elementary schools in Incheon, Republic of Korea. Participants had been studying English formally for an average of 3.7 years.

### Interventions and comparisons

Prior to the study, participants had been allocated by their schools to attend classes that were taught either by bilingual Korean/English teachers or by ‘native’ English-speaking teachers. In classes with bilingual teachers, the authors report that the majority of instruction was conducted in Korean, but that teachers often swapped between English and Korean (codeswitched) for a variety of pedagogic and administrative purposes. In classes with native English teachers, this approach was unavailable as the teachers were said to speak “little or no Korean” (Lee and Macaro 2013:887). These two conditions were described respectively as ‘codeswitching condition’ and ‘English-only condition’.

The authors took advantage of the “naturally occurring” (Lee and Macaro 2013:890) codeswitching and English-only conditions in the schools to compare the effects of teachers’ deliberate use of L1 to help teach a set of ‘target’ English words with use of only English for the same purpose. After selecting classes to participate based on their similarity in terms of the teaching approaches used by the teachers, the teachers’ professional experience, class sizes, ages, socio-economic backgrounds, and English proficiency of the students, the interventions were delivered as follows. Participants were given five minutes to skim read the text, then the teacher read through it with the class, stopping to discuss content and ask comprehension questions in English. When lack of vocabulary knowledge led to a breakdown in comprehension, the teacher clarified the meaning of the unknown

word, using Korean in the codeswitching condition and English in the English-only condition. The entire process was conducted on four occasions, with four different texts.

### Outcomes

Outcomes were assessed using a researcher-adapted version of the Vocabulary Knowledge Scale test (Paribakht and Wesche 1997). Participants' receptive recall of the target words was assessed by asking them to define each word, either in English or Korean. Then, their receptive recognition was assessed by asking them to select a definition from four possibilities for each word. This was then reassessed after a period of three weeks during which participants' education in English continued as usual.

### Results

Based on a one-way ANCOVA the authors report a statistically significant effect in favour of the codeswitching condition for recall and recognition of the target words, both on immediate post-tests and delayed post-tests, after controlling for baseline scores.

### Authors' conclusions

Lee and Macaro conclude that providing definitions in L1 results in better recall knowledge than using only English. In addition, they note that considerably less of the lesson time was taken up with clarifying unknown words in the codeswitching condition than in the English-only condition.

**Table 3.13** Risk of bias assessment for Study 8

<b>Selection Bias</b>	<b>Study Design</b>	<b>Confounders</b>	<b>Blinding</b>	<b>Data Collection Method</b>	<b>Withdrawals and Dropout</b>
Moderate:	Moderate:	High:	Low:	Moderate:	High:
Sample selected based on authors view that they were "representative" (p 890)	Two group pre and post	Non-random allocation	Assessor blind to group allocation, participants unaware of the research question	Researcher-adapted version of a validated test.	Not reported

## *9. Harnessing linguistic diversity in polylingual British curriculum schools. Do L1 mediated home learning tasks improve learning outcomes for bilingual children? A randomised trial*

**Reference:** Chalmers, H. (2014). *Harnessing linguistic diversity in polylingual British curriculum schools. Do L1 mediated home learning tasks improve learning outcomes for bilingual children? A randomised trial*. Masters Dissertation. Oxford Brookes University.

### **Research Question**

“Does the language with which children discuss the content of a reading book at home with family members have an effect on their subsequent performance on standardised assessments of written English for a written retelling of the same story?” (p18).

### **Participants**

Thirty-six children classified as EAL (English as an Additional Language) in Years 1 and 2 (ages 5 to 7 years) at one state funded primary school in Oxford, UK. Between them, participants represented 12 different L1s and a variety of ethnic backgrounds. Nine participants did not complete the intervention and were not assessed.

### **Interventions and comparisons**

Giving participants the opportunity to preview the contents of a storybook in their L1 was compared with previewing it in English only. The study responded to UK Department for Education advice to have EAL children discuss the content of a storybook in their L1 with a bilingual teaching assistant before carrying out work based on it in English (Bourne 2002). Due to the difficulty in finding teaching assistants proficient in the 12 languages represented among the participants, parents were asked to take on this role with their child. Children were randomly allocated to either an L1 group or an English-only group. Each participant was given a copy of a picture book with a clear narrative, but few words, and asked to read and discuss it with their parents over the course of 10 days. To facilitate this, parents were given a set of discussion prompts and guiding questions based on the kind of questions routinely used by UK teachers to promote engagement with texts and deeper understanding of them. The English-only group were given these prompts and questions in English, the L1 group had versions of them translated into their L1.

### **Outcomes**



Following the at-home intervention, the extent of the participants' understanding of the text was assessed by requiring them to produce a written retelling of it. The quality of the written retelling was scored using English National Curriculum Levels for English literacy and WIDA English Language Proficiency Performance Definitions.

### Results

Mean scores in all outcome measures were very slightly higher for participants in the English-only group, but these differences were not statistically significant.

### Authors' conclusions

Chalmers describes the study as inconclusive and suggests that if teachers in UK schools are to be confident that investing resources into acting on advice to use L1s is worthwhile, more and larger intervention studies are essential.

**Table 3.14** Risk of bias assessment for Study 9

Selection Bias	Study Design	Confounders	Blinding	Data Collection Method	Withdrawals and Dropout
Moderate:	Low:	Low:	Moderate:	Moderate:	Moderate:
All EAL pupils in the target age phase invited.	Randomised trial	Random allocation	Assessor blind to group allocation. Participants aware of the research question.	Validity and reliability of WIDA tool established, but NC levels problematic.	25% dropout

### Comment

This Master's dissertation is the only study of those included in this review to have been conducted in the UK and to have focused on a group of children who do not all share the same L1.

## 10. *The Effects of Using L1 Translation on Young Learners' Foreign Language Vocabulary Learning.*

**Reference:** Camó A.C. and Ballester, E.P. (2015). 'The Effects of Using L1 Translation on Young Learners' Foreign Language Vocabulary Learning'. *ELIA*, 15, 109–134.

### **Research Question**

“Does the use of L1 translation promote short and long-term vocabulary retention?” (p111).

### **Participants**

Forty children in Grade 5 (ages 10 – 11) of a Catalan primary school in Sant Hilari Sacalm, Spain. Sixteen participants were boys and 24 were girls. Most participants (70%) were speakers of Catalan, or both Catalan and Spanish in their homes. Some spoke only Spanish in their homes (9.5%). Some were of Moroccan origin and spoke Arabic in their homes (15%). One child was a speaker of Ukrainian and one child spoke both Portuguese and Catalan in his home. All participants had studied in the Catalan school system since the age of five, and were considered by the authors to be fluent speakers of Catalan. All had studied English as a Foreign Language at school since the age of five. Some had received additional English instruction outside school. The extent and nature of this additional instruction varied enormously.

### **Interventions and comparisons**

Participants were divided into two comparison groups. The authors do not report how participants were allocated. The intervention was delivered to each group as a whole. Participants were shown an English language animated version of the *Tale of Peter Rabbit*. When each of 20 preselected items of English vocabulary were encountered during the telling of the story, the teacher stopped the video and showed the participants a flash card on which the target word was written with an illustration of the item to which the word referred. In the L1 condition the teacher also told the participants the Catalan translation of the target word. In the English-only condition the word was not translated. The entire process was carried out twice for each group. Sessions were three days apart.

### **Outcomes**

Receptive knowledge of the target words was assessed through a multiple-choice format test, at three time points. The first test was administered immediately following the final teaching session. The second test was administered four days after that. Finally, a third assessment was administered one month after the final teaching session.

## Results

No statistically significant differences in scores between the two groups were detected on the immediate post-test. On the second post-test, scores were highly statistically significantly greater ( $P < .001$ ) in the L1 group. On the third post-test, the scores of the L1 groups remained highly statistically significantly greater than the scores of the English-only group.

## Authors' conclusions

The authors conclude that “although the use of the L1 is often neglected in the foreign language classroom, the present study suggests that the mother tongue can be used as a beneficial rather than a detrimental tool to promote foreign language vocabulary learning” (p127).

**Table 3.15** Risk of bias assessment for Study 10

<b>Selection Bias</b>	<b>Study Design</b>	<b>Confounders</b>	<b>Blinding</b>	<b>Data Collection Method</b>	<b>Withdrawals and Dropout</b>
Moderate:	Low:	High:	Moderate:	High:	High:
All children were learners of EFL, but selection criteria not reported.	Non-randomised comparison	Allocation schedule not described.	Not reported	Not reported.	Not reported, and discrepancies in these data between the two reports of this study.

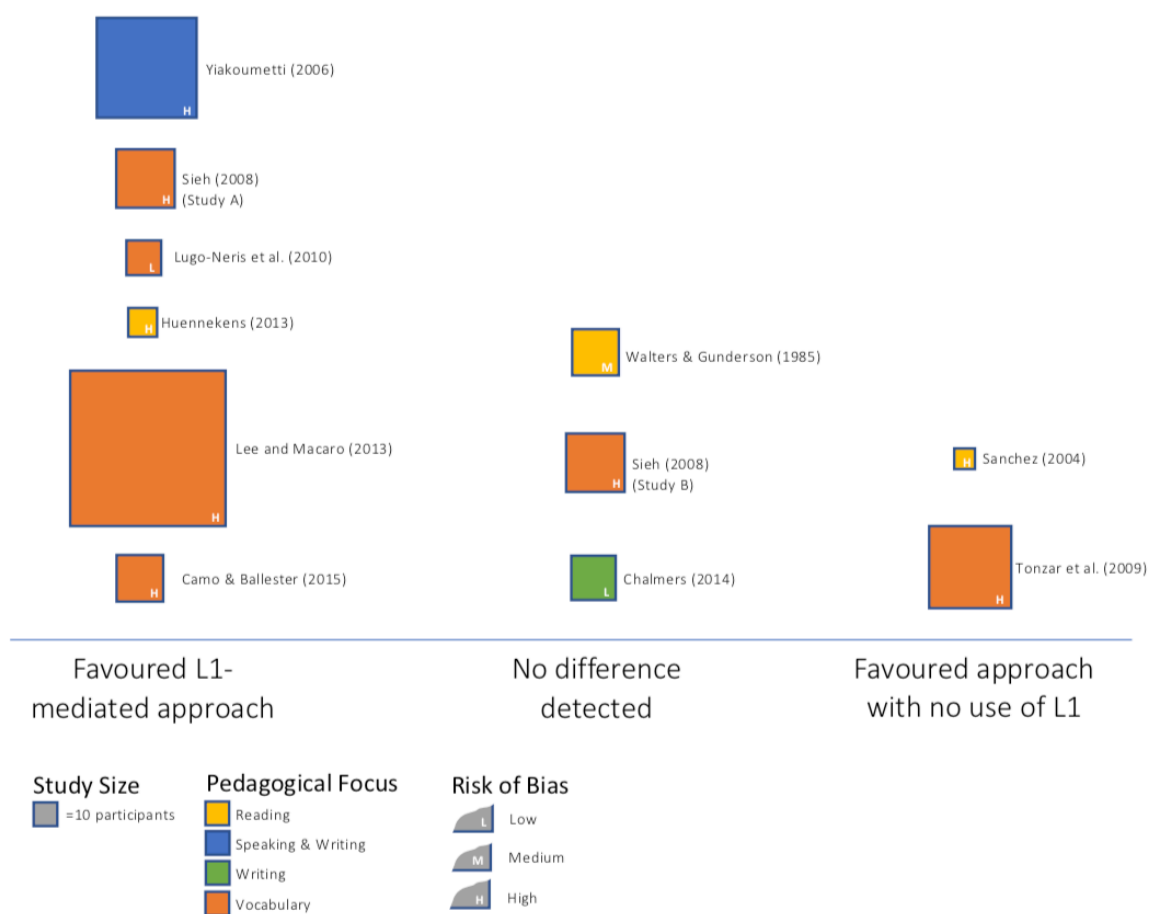
## Comment

This study also examined the reaction times by participants when they responded to a binary-choice receptive test of the target vocabulary. Participants who had been taught using the L1-mediated intervention were statistically significantly faster than those taught using the English-only intervention. The authors suggest that this is because “during the early stages of foreign language learning the FL [Foreign Language] lexis seems to be most likely organised in subordination to the L1 mental lexicon” (p127). That is, going from the L2 word to the mental concept via conceptual pathways already laid down in the L1 is more efficient than having to create new conceptual pathways in L2, at least in the early stages of L2 learning.

The study was based on a Bachelor's dissertation prepared by the first author. The published report of the study included a link to the full dissertation, which provided more detail about how the study was conducted. Both documents have been used to compile this summary. In the published document the authors state that the number of participants was 34, not 40 as reported in the original dissertation. As no account of dropout was given in either report, I have used the data from the original dissertation. Attempts to contact the author for clarification were unsuccessful.

### 3.5.7 Synthesis of results and bottom-line findings

The small number of studies that met the inclusion criteria, the variety of different outcome measures used in them, and the lack of data needed for confident calculations of effect sizes means that any attempt at statistical synthesis would risk being misleading, so I have not attempted this. However, as an indication of overall findings, Figure 3.3 represents the bottom line conclusions of each study as a function of risk of bias assessment. As can be seen, there are six studies that found L1 to improve outcomes, but five of these have a high risk of bias. Of the two studies with low risk of bias, one did not detect a difference between outcomes in the L1 and L2 groups (as did one study with ‘moderate’ risk of bias), and one found that when participants’ learning was mediated through their L1 they did better than when it was mediated only through English. Two ‘high’ risk of bias studies found that participants did better when their L1 was not used. Figure 3.3 helps to illustrate that, while findings appear to be equivocal overall, there is cause to be optimistic about using L1-mediated interventions to support the teaching of L2 vocabulary in primary schools. This cautious interpretation must be tempered by the lack of rigorous designs used in this body of research and incomplete reporting of these studies.



**Figure 3.3** Included studies by direction of outcome, size, topic and risk of bias assessment.

## 3.6 Discussion

This systematic review set out to describe the extent and nature of empirical research designed to assess the effects of teaching multilingual primary and pre-primary aged children using approaches that incorporate their L1 on their subsequent performance in academic and/or linguistic tasks conducted in their L2. The initial search of thirteen bibliographic databases, and studies identified by other means, returned titles and abstracts for 5,949 potentially relevant reports. Abstracts were screened for eligibility for the review based on eight inclusion/exclusion criteria, resulting in 210 reports that could not be excluded based on the information provided. I was able to locate full-text reports for 201 of these. The full-text reports were read in depth to determine their suitability for inclusion. This led to the exclusion of 200 reports because they did not meet at least one of the inclusion criteria. This left ten studies that met all of the inclusion criteria and were included in the final review.

The primary purpose of the review was to determine if there is sufficient evidence, of adequate trustworthiness, to allow teachers of primary and pre-primary multilingual school-children to make informed choices about whether and how their students' L1s should be incorporated into teaching approaches in a primarily L2 environment.

Despite the prevalence of advice to teachers, asserting that pedagogical use of students' L1s confers some sort of educative advantage on them (*e.g.* Bourne 2002, Butzkamm 2003, Celic and Seltzer 2013), and the equally common prohibition of L1s by schools or individual teachers (*e.g.* Kajonkiet International School 2014, Pattison High School 2017, Edstrom 2006) this review reveals that insufficient evidence (and a paucity of low-risk-of-bias studies) exists to support confident conclusions either way.

However, despite having revealed a paucity of research that would allow teachers to make confident decisions about use of L1 with their multilingual students, the review allows some inferences about which approaches may be promising, and therefore deserving of further research. A discussion of these inferences follows.

### 3.6.1 Pedagogic focus

#### 3.6.1.1 Vocabulary

It is noteworthy that, in such a small body of potentially relevant research, more than half of the studies that comprise it focus on vocabulary. The first of these (Walters and Gunderson 1985) had an incidental relationship to vocabulary in that the study used a standardised reading assessment, which incorporated a measure of vocabulary knowledge, to assess the effects of listening to stories

read in the L1. They used a robust design and concluded that being read to in L1 did not negatively affect understanding in L2. In five other studies, however, there was an explicit focus on the use of L1 to teach L2 vocabulary. Sieh (2008), Tonzar et al. (2009), Lugo-Neris et al. (2010), Lee and Macaro (2013), and Camó and Ballester (2015) all compared approaches that made deliberate use of L1 explicitly to teach L2 vocabulary with similar approaches that did not use L1.

Substantial evidence suggests that vocabulary knowledge is a strong predictor of reading comprehension in both monolingual and multilingual readers (*e.g.* Nation and Snowling 1998, 2004) and L2 vocabulary knowledge is an important predictor of communicative proficiency across the four domains of reading, writing, speaking and listening for individuals learning a new language (Milton 2013). In addition, EAL learners tend to lack the vocabulary knowledge needed to engage fully with their school curricula (*e.g.* Murphy 2014, Smith and Murphy 2015, Burgoyne, Whiteley and Hutchinson 2011). In their systematic review of intervention research on English language and literacy development for EAL learners, Murphy and Unthiah (2015) identified a similar pattern to this review. A majority of the 29 studies that met their review criteria focused on some aspect of teaching vocabulary. A large proportion of those focused on comparisons of different ways to teach academic vocabulary, specifically. Academic vocabulary - words that are unlikely to be encountered in day to day conversation but which are common in the kinds of discourse encountered in school - are especially important to teach, precisely because of their scarcity outside academic contexts (Cummins 2011). Given our knowledge about the importance of explicit vocabulary teaching for EAL learners, and the assumption that L1 is a useful tool in L2 learning, attempting to find ways to enhance vocabulary acquisition in EAL learners by using their L1 would appear to be well warranted.

Of the five studies in this systematic review that compared ways to explicitly teach vocabulary using either L1 or not using L1, three found statistically significantly higher average scores on measures of vocabulary knowledge when teaching had incorporated participants' L1s (Lugo-Neris et al. 2010, Lee and Macaro 2013, Camó and Ballester 2015). One study (Sieh 2008) found mixed results, with advantages to using the L1 detected in one iteration of the experiment but no such advantage found in the other iteration. The fifth study found the opposite: using L1 was associated with statistically significantly lower scores on tests of vocabulary knowledge (Tonzar et al. 2009). An interesting difference between the four studies that found an advantage of using L1 and the one study that found a disadvantage is the extent to which each approach reflected what might be considered typical classroom practice. In the studies by Sieh, Lugo-Neris, Lee and Macaro, and Camó and Ballester the intervention was delivered in fairly typical teaching contexts. In all of these studies, the intervention was delivered as part of a whole class reading activity, stopping at key items of vocabulary to explain them either using the L1 or the L2. Lugo-Neris et al.'s participants were taken

out of class to work in small groups with an adult (not dissimilar to the way Teaching Assistants in schools are often deployed to provide targeted support to individuals and groups of children). Participants were read a book by the adult, then, on encountering problematic vocabulary items, their meanings were explained by the adult in either Spanish (the participants' L1) or in English. Tonzar et al.'s intervention, on the other hand, while delivered in real schools, could legitimately be described as a laboratory study. Participants were shown a total of 40 L2 words paired with either their L1 translations or pictures of them, at a rate of one pair every eight seconds. This is an unusual way to teach vocabulary. The way they measured vocabulary knowledge was just as atypical. Participants were presented with either an L1 word or a picture of one of the words they had 'learned'. They were then given twelve seconds to write down its L2 translation. This was repeated until all 40 stimuli had been presented. It is worth considering whether this is a helpful demonstration of what it means to *know* an item of vocabulary (see Nation 2001) and, therefore, if it is helpful to teachers of EAL learners. The contexts in which the comparisons were made in these studies are not *a priori* reasons to dismiss or accept their findings. Nonetheless, the extent to which they differ from routine classroom practice is an important factor in assessing the utility of their findings for teachers wishing to make evidence informed decisions about their practice.

To summarise the above discussion, the warrant for vocabulary learning as an object of study is sound and, based on the small amount of relevant research that I have located, using L1 to teach vocabulary in naturalistic classroom contexts holds some promise. The teachers in these studies were responsible for providing the L1 translations, and were well positioned to do so as they were proficient in the students' L1 as well as the target language. While this does not address the logistical problem of L1 use by teachers in linguistically diverse classrooms, it is possible that providing pre-recorded translations of key vocabulary, written L1 glosses, bilingual dictionaries, or machine translators for multilingual children in linguistically diverse classes would serve the same purpose. Empirical evaluation of strategies such as these is worth consideration.

### 3.6.1.2 General reading proficiency

Two studies assessed the effects of using L1 to support activities around reading. Two of these assessed the effects of such activities on measures of comprehension of written text, as demonstrated in English. Sánchez (2004) gave eight Spanish L1 children Spanish language translations of English story books to read with their parents, then assessed the children's comprehension of the original English language versions as well as the speed and accuracy with which they read them (their 'oral reading rate'). Using a multiple baseline, interrupted time series design, these measures were taken on several occasions over time to generate slopes to describe the trajectory of their development. The slope values before and after the intervention was introduced



were compared to help make casual inferences about the intervention and any effects it may have had. Sloped values before and after the introduction of the L1 intervention did not differ statistically significantly, and in many cases slopes decreased in magnitude, suggesting that participants were making better progress when they were being taught by their teacher's usual approach.

The other study to assess the effects of L1 use on a reading-related skill (Huennekens 2013) did so by implementing one-to-one 'dialogic reading' in Spanish then assessed the phonological awareness and alphabet knowledge in English of fifteen participants. Huennekens found a statistically significant improvement in scores on these measures before and after the intervention, and suggested that this difference was caused by the intervention. However, because she used a single group pre/post design, there is insufficient evidence to infer what the magnitude and direction of that difference might have been had the participants not been exposed to the intervention.

In summary, the available evidence on L1 use and its relationship to reading comprehension and knowledge of letter sound relationships is not sufficiently clear to allow teachers to make informed decisions about whether to try to implement it in their classrooms.

### 3.6.1.3 Oral and written proficiency in the L2

Yiakoumetti's (2006) study assessed the effect of explicit comparison of L1 and L2 forms on the oral and written production of the L2. Here, students were encouraged to notice the differences between a variety of aspects of their L1 and L2, such as differences in syntax and morphology. By making this explicit, it appears that the intervention had a corrective effect and reduced the frequency with which participants made errors in their L2 that could be attributed to interference by their L1. This was true for both their oral and written language. This approach relies on the teacher possessing sufficient knowledge about the pupils' L1 and its relationship to L2. Therefore, the approach has limited practical value for teachers with a linguistically diverse group of pupils. Notwithstanding my earlier comment about the difference between a language and a dialect, the two languages of the participants in Yiakoumetti's study, Cypriot Dialect and Standard Modern Greek, are linguistically close. As a result, the findings of this comparison may be of limited value when L1s are linguistically distant from the target language, such as Urdu or Bangla in English schools.

Chalmers (2014) gave parents of EAL children L1 question prompts to facilitate exploration of the events in a picture book with their children. Participants' comprehension of the story was then assessed by asking them to re-write it in English. The assessment focused on general writing skills, which included the cohesiveness and completeness of the text. The results of participants who had used the L1 to explore the story were compared with those of children whose parents had been

given the question prompts in English. No statistically significant differences in outcomes between the comparison conditions were detected.

### 3.6.2 Location and target population

An important finding from this systematic review is what it tells us about the contexts in which L1-mediated teaching approaches have been exposed to empirical assessment.

Three of the ten studies were conducted in the USA. The educational and linguistic context for multilingual learners in the USA is quite different to that in the UK. In particular, the distribution of languages across the population of non-English-L1 speakers is more amenable to making use of L1 in the classroom in the USA than it is in the UK. In the USA, 62% of non-English-L1 speakers speak Spanish or a Spanish-lexified Creole (Ryan 2013). The next largest minority language is Chinese, which is represented by a much smaller proportion of the non-English-L1 population, only 4.8%. The other languages spoken in the USA are represented by still smaller proportions of the whole. Contrast this to England and Wales where the largest minority language (Polish) is spoken by a comparatively small 13% of the non-English-L1 population. This is followed by Panjabi, Urdu, Bengali, and Gujarati, each of which is spoken by between 5% and 6.5% of the non-English L1 population, then Arabic, French, Chinese, Portuguese and Spanish, each of which represents between 2.8% and 3.8% of the languages spoken by non-English-L1 speakers (Office for National Statistics 2013). In the USA, the overwhelming majority of non-English-L1 speakers using Spanish makes drawing on L1 as an explicit teaching tool comparatively straightforward. Slavin and Cheung (2005) recognise the importance of this in their systematic review of bilingual programmes, describing a “critical mass” (Slavin and Cheung 2005:250) of Spanish dominant students as a key factor in making Spanish-English bilingual schools viable. In the studies conducted with Spanish-L1 participants in the USA, teachers are more likely to know some Spanish (or, indeed, be Spanish-English bilinguals themselves), students can work together using Spanish, Spanish language resources are likely to be readily available, and there is a degree of socio-cultural similarity among students that may have a facilitative effect on learning. This general point is also applicable to the studies in Canada, Italy, Cyprus, Taiwan, Korea, and Spain. The Canadian study was conducted in “a predominantly Cantonese speaking school” (Walters and Gunderson 1985:67) and the Italian, Taiwanese, Korean, and Spanish studies were conducted in English as a Foreign Language (EFL) classes, in which the vast majority of students will have shared the same L1 (Italian, Mandarin, Korean or Catalan). In the study in Cyprus students were all speakers of Cypriot Dialect at home, as were their teachers. In the linguistically more heterogeneous UK, none of these things can be assumed, and meeting the challenge of using L1 with EAL learners is commensurately much more complex. The only study to evaluate using L1 as a teaching tool in linguistically diverse classrooms (Chalmers 2014) was inconclusive.

In summary, making generalisations, or extrapolating lessons for practice in other contexts from the studies described here must be done with caution.

### 3.6.3 Conclusions based on the quality of evidence (best evidence synthesis)

This review adopts a best evidence synthesis approach (Slavin 1986) to acknowledge that designs considered to be the most rigorous in reducing bias are uncommon in educational research (though this is slowly improving (Connolly 2015)), and to anticipate that few studies of the highest methodological rigour for making causal inferences might be located for this review. Individual elements that contribute to the overall trustworthiness of each study have been described in the results section. In this section I will discuss the combined findings of studies of broadly comparable trustworthiness, based on study design.

#### 3.6.3.1 Randomised trials

Three studies used the design considered to be the most trustworthy for informing causal inferences (Campbell and Stanley 1963), by randomly allocating participants to alternative interventions. These were Walters and Gunderson (1985), Lugo-Neris et al. (2010), and Chalmers (2014). Walters and Gunderson and Chalmers did not detect statistically significant differences in outcomes between the approaches they compared. Lugo-Neris et al. found that providing L1 explanations of L2 vocabulary items led to better expressive and receptive knowledge of those vocabulary items.

Overall, the randomised trials that compare L1-mediated teaching approaches with alternatives in this body of evidence do not provide a clear picture of the differential effects of the L1-mediated and English-only teaching approaches they investigated.

#### 3.6.3.2. Non-randomised comparisons

Five studies assessed the effects of L1-mediated teaching approaches by comparing the outcomes of pre-existing groups that were allocated to alternative teaching approaches using a method other than random (or otherwise unbiased) allocation: Yiakoumetti (2006), Sieh (2008), Tonzar et al. (2009), Lee and Macaro (2013), and Camó and Ballester (2015). Yiakoumetti (2006) found that pupils taught to deliberately compare and contrast features of their L1 with their L2 subsequently made fewer mistakes in their L2. Sieh (2008), Lee and Macaro (2013), and Camó and Ballester (2015) found that using the L1 to define new items of vocabulary resulted in better knowledge of the target words. Tonzar et al. found that groups that had been taught vocabulary using L1 did less well at remembering those words than groups taught using an approach that did not use L1. The way in which participants in these studies were allocated to comparison conditions was not clear from the reports, though four studies made some effort to demonstrate that groups were similar on some baseline characteristics, such as scores on baseline assessments and socioeconomic status

(Yiakoumetti 2006, Sieh 2008, Lee and Macaro 2013, Camó and Ballester 2015). As a result, while it is possible to assert similarity between groups on some of the known characteristics of the participants, systematic error from unmeasured characteristics cannot be ruled out.

Overall, assessment of L1-mediated teaching approaches that used non-random allocation to comparison groups suggests that, on balance, using L1 has a positive effect on L2 outcomes. If one considers the studies that focused on the same outcome, vocabulary knowledge, the findings of these four suggest a warrant for evaluating L1-mediated vocabulary teaching using more robust experimental designs. However, as they stand, the findings of these studies must be treated with more caution than those derived from studies using random (or otherwise unbiased) allocation to comparison groups.

#### 3.6.3.4 Designs with no comparison group

One study (Sánchez 2004) used a multiple-baseline, interrupted time series design. This compared series of scores on reading comprehension and oral reading rate for individual participants in time periods before the introduction of the intervention and in time periods while the intervention was being delivered. It found no clear pattern to suggest that the intervention made a difference to the outcomes, and in many cases data suggested that it might have made things worse. Another study used a single group pre/post design (Huennekens 2013). It found that phonological awareness and alphabet knowledge improved over time for that group.

Overall, one inconclusive study and one study for which it is impossible to disentangle intervention effects from other possible effects (such as changes in the students or to intervention delivery over time) do not provide a clear picture of the effects of introducing an L1-mediated teaching approach on the outcomes studied.

#### 3.6.3.5 Summary

When studies eligible for inclusion in the review are classified by the risk of bias associated with their designs, and considered for narrative synthesis on those terms, no clear picture about the effects of L1-mediated teaching approaches emerges from the most robust or the least robust. However, given the glimpse of promise regarding vocabulary teaching derived from non-randomised comparisons (which might be thought of as occupying a space somewhere between the most and least robust), the combined findings of these studies warrants further exploration using studies better designed to reduce non-random error.

## 3.7 Conclusions

Primary school teachers are torn between official advice to use their students' L1s as a pedagogic tool on the one hand (e.g. Bourne 2002), and formal or informal prohibition of such approaches in their schools on the other (e.g. Pattision High School 2017, Edstrom 2008). This review set out to address this implicit collective uncertainty by investigating the following questions:

RQ1. What is the nature and extent of empirical research on the effects of using L1-mediated teaching approaches in L2 settings with primary and pre-school-aged language learners, on outcomes relating to either L2 proficiency or academic attainment, or both?

And

RQ 2. What, if anything, can be concluded about the effects of L1-mediated teaching approaches in L2 primary school settings?

Responses to these questions are summarised in the following sections.

### 3.7.1 The nature of the research

#### 3.7.1.1 Skills focus

In the main, the skills focus in the literature was related to L2 literacy. Only one study addressed skills specifically associated with speaking. No studies addressed outcomes associated directly with listening and understanding.

Four studies investigated the effects of L1-mediated teaching approaches on one or more skills related to reading: reading comprehension, oral reading rate, vocabulary knowledge, phonological awareness, and letter knowledge. One study used reading as the vehicle for using the L1, but assessed outcomes associated with writing proficiency.

Explicit vocabulary teaching formed the focus for five studies, all of which compared using the L1 to explain the meanings of unfamiliar words with using alternatives that used only the L2, or only picture cues. In four of these studies, vocabulary teaching was contextualised through reading activities.

One study assessed the effects of using the L1 as a pedagogical tool on oral and written proficiency in the L2.

All but one study was conducted in 'naturalistic' contexts. That is to say, the interventions were carried out in schools or in the home and reflected the kinds of activities that take place in those

places, as a matter of course. One study is better described as a ‘laboratory’ study. While it took place in schools, the intervention did not reflect typical classroom practices.

### 3.7.1.2 Location and languages

Four of the ten studies were conducted in North America (three in the USA and one in Canada). Four were conducted in Europe (one each in Cyprus, Italy, England, and Spain). Two were conducted in East Asia (Taiwan and South Korea). In one study, the participants represented a variety of L1s. In all other studies the participants shared the same L1, either the main language of the region in which the study was conducted (Mandarin, Korean, Italian, Cypriot, Catalan), or that of significant linguistic minority in the country (Spanish, Cantonese).

### 3.7.1.3 Methods

Methodologically, the majority of studies (eight of ten) used designs robust enough to allow relatively confident causal inferences to be made, though only three of these unequivocally allocated participants to interventions using unbiased methods. These studies taken together do not clarify whether L1-mediated approaches are helpful or not (though there appears to be promise in using the L1 to teach L2 vocabulary). The remaining studies used designs that are less robust, which, when taken together do not give a clear indication of the effects of using pupils’ L1s either.

## 3.7.2 Extent of the research

The extent of empirical research that assesses the effects of L1-mediated teaching approaches in non-bilingual primary and pre-primary schools is extremely small; only ten relevant studies were located. Of these, about a third were located in grey literature, in this case Master’s and Doctoral theses. Only seven were published in peer reviewed journals in the thirty-five year period covered by this review (and one of these was a write up of a Bachelor’s dissertation). The ubiquity of advice to use the L1 as a pedagogical tool for multilingual learners in primary schools is extremely hard to reconcile with the apparent lack of interest among the SLA community in evaluating related approaches and publishing them in the peer reviewed literature.

## 3.7.3 Implications for practice

Teachers in non-bilingual primary and pre-primary schools who are considering investing in ways to use their multilingual pupils’ L1s as a teaching and learning tool in the belief that it will raise those pupils’ L2 proficiency or their academic attainment must do so cautiously. Equally, policy makers in primary schools who are considering imposing ‘English Only’ rules or other prohibitions of pupils’ L1 use in their schools must act with similar caution. There is an extremely small body of research comparing L1-mediated teaching approaches with alternatives in such contexts, and, taken as a whole, the results of this body of research are equivocal.

There appears to be promise in using children's L1s to explain the meanings of L2 vocabulary items. The literature suggests that this approach might be most effective when it incorporates discussion of the words in a meaningful context, such as during reading. This has only been assessed in linguistically homogeneous groups of students, and so the implications for teachers of students who represent a variety of different L1s are not clear.

There may also be promise in helping multilingual children to identify similarities and differences between their L1 and L2 as a means to improve production in the L2. Though, implementing this requires that teachers possess considerable knowledge about the linguistic structures of both languages, and thus provides a significant challenge to teachers of linguistically diverse classes.

### 3.7.4 Implications for future research

In order for primary and pre-primary school teachers to have clear evidence upon which to base their understanding of the effects of using L1-mediated teaching approaches in L2 settings, much more intervention research needs to be conducted.

Using L1 to teach vocabulary may be helpful. More comparisons of L1-mediated vocabulary teaching with approaches that use only L2 would help to determine whether the positive findings of the three small studies described here are more widely applicable.

Studies that assess the effects of using L1 on proficiency in writing, speaking and listening are notable by their relative absence from the literature. Research that assesses, for example, the effects of asking children to draft written compositions in L1 before re-writing them in L2 would be valuable additions to the evidence base.

The overwhelming majority of studies in this review were conducted in contexts where all students shared the same L1. It is imperative that more research is conducted in contexts where a variety of L1s are represented. Linguistically diverse classrooms, particularly in Europe where transmigration of linguistically heterogeneous populations is common, are increasingly the norm. Research to assess ways to meet the language needs of a linguistically diverse classroom should include a combination of efficacy trials that assess whether the principle of L1-mediated teaching is sound, and process evaluations to help understand the delivery of interventions to linguistically diverse groups of students.

### 3.7.6 Limitations of this systematic review

I have made every effort to locate as much relevant research as possible given the resources available to me while preparing this review. However, inevitably some reports may have been missed.

#### 3.7.6.1 Electronic searching, hand searching, and language of publication

All searches were conducted initially using electronic resources, then reference lists of included reports were scanned for additional possibly relevant studies. However, no hand searching of paper journals was done. In addition, although efforts were made to include studies published in languages other than English, this was only possible if they had English language abstracts. Clearly, searches that use only English search terms will return only records of foreign language studies for which authors have taken the trouble to translate their abstracts into English. While this might be common for peer reviewed literature it is less likely to be true for the grey literature. Also, while I did contact some authors of eligible and potentially eligible studies to ask for clarification about elements of their reports, and while doing so asked them if they knew of any other studies that may meet my inclusion criteria, I did not canvass for suggestions from the academics in the field more generally. Therefore, my efforts to include as much of the relevant information on L1-mediated learning in L2 settings as possible can only be said to go so far. I must concede that I may have missed studies published in languages other than English that might make important contributions to understanding the totality of the evidence.

#### 3.7.6.2 Grey literature and publication bias

Locating grey literature was not limited to issues of language of publication. It is possible that there are relevant studies written in English that have not been catalogued in the online bibliographic databases, but which would have nonetheless added important information to this review. In addition, publication bias is likely to be responsible for making it difficult to locate the totality of the relevant evidence. Publication bias (where so called 'null' and 'negative' findings are systematically underrepresented in the published literature compared to so called 'positive' findings) affects all areas of academic research, and research on bilingualism is no different (De Bruin et al. 2014).

#### 3.7.6.3 Quality assurance

I was able to recruit help to double screen a portion of the studies included in the second screening phase (studies that could not be excluded on the basis of their abstracts alone), but I was unable to recruit help for the initial abstract screening, or to double data extract from the studies that were ultimately included in the review. In any update of this review, a second reviewer must be recruited to help minimise potential bias in assessing the literature.



Finally, the original search for this review was conducted in April 2015. New evidence may have been generated in the intervening time. An update of this systematic review is necessary if it is to be considered for wider publication.

### 3.8. Informing the direction of the next stage of this doctoral project

The aim of this systematic review was to explore the extent and nature of the empirical research on L1 use in L2 contexts, with the expectation that this would reveal gaps in the evidence about these approaches and thus indicate fruitful avenues to pursue during the second part of this doctoral project. As is clear from the findings, the gaps are so large that one might reasonably 'stick a pin in the map' and go from there. Despite books filled with advice to teachers about using the L1 as a pedagogical tool (e.g. Celic and Seltzer 2013, Hesson, Seltzer and Woodley 2014, Espinosa, Ascenzi-Moreno and Vogel 2016), the vast majority of the approaches contained in them appear not to be based on empirical comparisons with alternatives. All, therefore, deserve to be rigorously evaluated in controlled comparisons and, thus, any of the strategies described in these books could form the focus of my intervention study.

#### 3.8.1 Deciding the focus of the intervention study

One general approach that looks promising from the few studies that have evaluated it, is the use of L1 to support the teaching of L2 vocabulary. The four studies included in my systematic review that found an advantage to supporting L2 vocabulary learning using the L1 operationalised the L1 by translating target vocabulary items into it and using it to provide supporting information about them. In these studies all participants spoke the same L1, as did their teachers. Teachers were able easily and efficiently to provide the L1 input required for the intervention. In my context this is not possible. The children who theoretically stand to benefit from this type of approach in English primary schools speak a variety of different L1s, and their teachers are unlikely to be sufficiently proficient in all (or any) of the L1s spoken by the pupils in their classes. Even if they were, pausing to translate and describe items of vocabulary in all of the L1s represented in a class would be extremely time consuming and disruptive to the flow of the lesson. Accepting that using the L1 to teach vocabulary would be an appropriate response to the findings of my systematic review, my challenge was to design and test an intervention that uses the L1 to support L2 vocabulary learning in a way that does not rely on the linguistic proficiency of the teacher.

In two of the books cited above, *Translanguaging: A CUNY-NYSIEB Guide for Educators* (Celic and Seltzer 2013) and *Translanguaging in Curriculum and Instruction Guide* (Hesson, Seltzer and Woodley

2014), teachers in linguistically diverse classes are advised to use a ‘four-box graphic organiser’ or ‘Frayer Model’ as a vocabulary teaching strategy amenable for use with pupils representing a variety of L1s. The four-box graphic organiser and the Frayer Model are types of concept map. Concept maps are blank frameworks that provide spaces to note and organise information about the item of vocabulary being learned. The teacher provides information about the target word and leads a discussion with the pupils around the word. The teacher then demonstrates adding information about the word to the concept map, which the pupils then continue to do independently. The nature of this information varies depending on the type of concept map, but usually includes information such as the word itself, a pupil-friendly definition of the word, examples of things represented by the word (if learning the word ‘stationery’, for example, the pupils or teacher might include ‘pen’, ‘pencil’, and ‘ruler’), and sentences contextualising the word in use. The above-mentioned guides suggest that teachers use Google Translate to prepare L1 exemplars to use when they model adding information to the concept map, and encourage their pupils to use their L1 to discuss the target word among themselves, and to add additional information to the concept map.

Concept maps of this sort have been shown to be effective for developing vocabulary knowledge with monolingual students in their own languages (*e.g.* Peters 1974, Monroe 1997), and with second language learners using only the L2 (Nahampun and Sibarani 2014). I am not aware of any research that has evaluated the effects of concept maps on L2 learning when their use has been mediated through the students’ L1. Evaluating this approach, therefore, is an appropriate response to the findings of my review.

### 3.8.2 Warrant for the intervention study

Basing my primary intervention study on a comparison of an L1-mediated concept map strategy with an English-only version of the strategy on vocabulary learning in the L2, is well warranted for the following reasons:

1. It responds to indications from my systematic review that vocabulary learning is enhanced when the L1 is used, but extends the exploration of this to contexts in which teachers do not speak the L1s of their pupils, and are responsible for the education of linguistically diverse groups of children.
2. It adapts an approach that is - in monolingual form - already widely used by teachers, and thus allows for a meaningful comparison to be made.
3. Advice to use an L1-mediated version of the concept map is already in circulation and may, therefore, be being adopted despite a lack of evidence of its efficacy. This study will

contribute much needed evidence to help teachers decide whether the approach is a worthwhile use of their resources.

The second of these points, that the L2 version of this strategy is supported by empirical evidence and already widely used, is particularly important as far as carrying out a meaningful comparison is concerned. It has long been a principle in health research that, if the results of comparisons of alternative interventions are to be meaningful, novel interventions should be compared to the best available alternative, rather than to nothing at all (*i.e.* a placebo) (World Medical Association 1975). Health professionals are not interested in whether a new intervention is better than no intervention, they are interested in how it compares to what they already do. Connelly et al. (2017) discuss this principle in relation to education research, noting that it is rarely possible or ethical to compare a novel educational intervention with the educational equivalent of a placebo (*i.e.* nothing at all). Instead, comparisons of novel educational interventions tend to be compared with a 'business as usual' control. That is to say, the control group carries on with whatever it would normally be doing while the intervention group engages with the novel intervention. This might appear to conform to the principle of a meaningful comparison. The problem is, however, that 'whatever the control group would normally be doing' is often unspecified and may not be at all related to the focus of the novel intervention. For the purposes of the comparison, therefore, it might as well be the equivalent of a placebo. Teachers do not want to know whether, for example, a new vocabulary intervention is better than something unspecified and possibly unrelated to vocabulary teaching, they want to know whether it is better than what they already do when they teach vocabulary.

The desirability of meaningful comparisons expressed through this principle is gaining ground in the discourse on educational research. This was recently evidenced by educationalist Dylan Wiliam, who, in a discussion on innovation in teaching, suggested that "Doing things that might work, while not doing things that are known to work is, in my view, unprofessional, like withholding effective treatments for illnesses in order to research others" (Wiliam 2018a). Instead, he said, we should "compare new approaches to the best (*i.e.*, most cost-effective) available approaches, rather than to just "business as usual."" (Wiliam 2018b). By choosing to compare the effects of an L1-mediated concept map with an alternative that enjoys empirical support and is already widely used, I adhere to this principle and will hopefully generate information that is meaningful to teachers.

### 3.8.3 Accommodating linguistic diversity

As I have mentioned, Celic and Seltzer (2013) suggest using Google Translate to provide written L1 input for all of the languages represented by students. However, to benefit from this approach participants must be literate in their L1. My target population - EAL learners in mainstream English

primary schools - includes individuals who speak their L1 or understand it when someone else speaks it, but who are unable to read or write it. I have concluded that to make the L1 intervention accessible to as many participants as possible, I should find a way to provide spoken L1 input for them. One way to do this is to create audio recordings of the information usually provided by the teacher when she introduces the word at the beginning of the concept map activity, translated into the L1s represented by her students.

### 3.8.4 Overview of the intervention study

To summarise, based on the findings of my systematic review, knowledge of existing (monolingual) approaches to vocabulary teaching, and recognition of recommended, yet apparently evidence-free L1-mediated versions of this approach, the intervention phase of this project aims to address whether using audio recorded pedagogical input in the L1s of multilingual learners in linguistically diverse mainstream English primary schools is superior to using only English. The intervention and the approach taken to assess its effects are described in detail in the next chapter.

# PHASE 2: INTERVENTION STUDY

# CHAPTER 4 METHODOLOGY

In this chapter the specific research questions that this study addressed are presented. The role of paradigms in informing approaches to educational research are discussed, and the method adopted to address the research questions is presented alongside the rationale for this choice. The characteristics of the schools and children chosen to participate in the study are described and justified. The chapter also describes how the intervention materials used in the study were developed, and it explains the procedure for the delivery of these interventions and how their effects were assessed. The chapter concludes with a description of the steps taken to meet the ethics requirements for research with human participants.

## 4.1 Aims and objectives of the research

At the macro-level this research aimed to provide much needed empirical evidence to inform our interpretation of the claims inherent in theoretical constructs such as linguistic interdependence (Cummins 1979) and translanguaging (García and Li Wei 2014). Proponents of both theories suggest that, due to the interconnectedness of languages in the minds of multilingual individuals, learning mediated in one language will transfer to performance in the other. Indeed, these theoretical positions have been interpreted by many to mean that learning through the L1 will result in improved L2 performance, compared to learning only through the L2 (Celic and Seltzer 2012, Chumak-Horbatsch, 2012). In essence, the veracity of such claims will be informed by the results of the study, either providing support for the hypotheses or facilitating a reconsideration of their application for pedagogy. This is taken up at the micro-level in the study, where theories of linguistic interdependence and translanguaging are applied in the classroom for a specific purpose. Thus, the specific aim of the intervention study was to evaluate the effects on recall and recognition of English language vocabulary when the L1s of students' in linguistically diverse classes are used as mediating tools. Taken together, the macro- and micro-level aims of the research thus direct us to consider carefully and empirically what can legitimately be claimed based on linguistic interdependence and translanguaging theories as they related to classroom practice.

## 4.1.2 Research questions

The intervention study addressed the following research questions:

1: What are the differential effects of using primary school EAL pupils' L1s to mediate English language vocabulary instruction, compared with using only English, on expressive and receptive knowledge of the target words?

2: Are there any statistically significant associations between the level of success on tests of expressive and receptive knowledge of the target words and participants' English proficiency, L1 proficiency, attitude to L1, length of residence in the UK, or time spent in English schooling?

## 4.1.3 Hypotheses

The working hypotheses of the study were as follows.

### **RQ1**

H<sub>1</sub>: Using the L1 of primary school pupils classified as EAL to mediate their learning of English language vocabulary is associated with different outcomes on tests of expressive and receptive knowledge of the target words, compared with using only English.

H<sub>0</sub>: The language with which English language vocabulary instruction is mediated for primary school pupils classified as EAL is not associated with differential outcomes on tests of expressive and receptive knowledge of the target words.

### **RQ2**

H<sub>1</sub>: Individual characteristics of participating pupils, such as proficiency in English and proficiency in L1, are meaningfully related to attainment in tests of expressive and receptive knowledge of the target words.

H<sub>0</sub>: Individual characteristics of participating pupils, such as proficiency in English and proficiency in L1, are not meaningfully related to attainment in tests of expressive and receptive knowledge of the target words.

## 4.2 Research paradigms, ontology and epistemology

In 2013 The Department for Education in the UK published a provocation paper by doctor, academic and popular science writer Ben Goldacre entitled *Building Evidence into Education* (Goldacre 2013). The paper argued that education was ripe for a fundamental metamorphosis into a truly evidence-informed profession, where intervention research, and in particular randomised trials, are conducted as a matter of routine to test theory, to make educational claims falsifiable, and ultimately to help increase the professional independence of teachers and improve outcomes for children. Goldacre's provocation could hardly have been more effective. It helped to catalyse sustained, vibrant and sometimes bloody exchanges about educational ontology and epistemology among teachers, academics and policy makers.

Goldacre's paper was met with howls of existential angst from some sections of the educational hegemony. Gary Thomas, a staple on the educational research methods shelves of graduate students, characterised Goldacre's argument that randomised trials should play a bigger part in informing education practice as "merely faith" (Thomas 2013a, no page). Sociologist Frank Furedi (2013, no page) seemed to suggest that teaching is uniquely resistant to exploration by experiment and that we should, therefore, "keep the scourge of scientism out of schools". Tropes arguing that "randomised controlled trials belong to a discredited view of science as positivism" (Cohen, Manion and Morrison 2007:278) or that "educational research [should] take the form of case studies rather than randomised controlled trials" (Elliot 2001:564) were recycled in the ensuing debates (Connolly, Keenan and Urbanska 2018).

By contrast, the first public response to Goldacre's paper from teachers was the establishment of ResearchED, "a grassroots, teacher led organisation" (ResearchED 2018a, no page) that inserted teachers directly into the conversation about the interface between their practice and the research that ought to inform it. 'Working out what works' was the organisation's strap line for its inaugural conference in 2013, alluding to their broad agreement with Goldacre that educational research could do better if it embraced randomised trials. Indeed, the founder of ResearchED cites Goldacre's paper as directly responsible for the establishment of the organisation (ResearchED 2018b). Part of the UK government's response was to designate the then recently established Educational Endowment Foundation (EEF) as its *What Works Centre for Education* (EEF 2018a). The EEF commissions, almost exclusively, randomised trials to evaluate educational interventions. In its six years of existence has been responsible for more government funded randomised trials in education than ever before (EEF 2018b). Trialists became emboldened to argue more forcefully the case for



randomised trials in education where historically, in the UK at least, they had been a largely ignored minority voice (Connolly et al. 2017).

All this is to illustrate a tendency towards tribalism in education research that, at its worst, frames ontological and epistemological paradigms as mutually exclusive academic silos standing in stoic irrelevance to one another. If you value randomised trials, says the orthodoxy, you are a positivist: broadly, the assumption that “only observable things can be included in valid knowledge, that theory is a compendium of empirically established facts [...] and that the act of trying to know ought to be carried out in such a way that the knower’s own value position is removed from the process” (Oakley 2000:31). If you prefer case studies you are considered an interpretivist: broadly, the assumption that no objective reality exists because reality is constructed uniquely by each individual, that to understand social phenomena we must immerse ourselves in the world we are studying, and that researchers inevitably bring their own unique reality to bear on how they interpret what they observe (Thomas 2013b). You must choose your silo and argue vociferously with your opposite numbers as to why they are wrong. Ann Oakley, doyen of social sciences research methodology, calls this the “paradigm wars” (2000:23). But as Oakley and others (Gorard 2013, J Thomas et al. 2004, Oliver 2015) have argued, this manufactured war has no basis in the reality of conducting educational research. Whatever silo a researcher has been encouraged to inhabit, when they do their research they adopt parts of the personae of positivists *and* interpretivists.

For many years, on the wall of my dad’s office was pinned a strip of paper with the badly photocopied *bon mot* “Seek not truth, be happy with .05” (he was a medical researcher). This sits firmly in the positivist camp. But note the admission that we can never know the truth (and must be content with .05, the statistical probability that what you have observed happened by chance). My dad’s pre-internet meme alluded to the idea that knowledge is always provisional and imperfect, and that there may be alternative explanations to why something happened. This is also reflected in the recognition by trialists that mere involvement in research can change participants’ behaviours through placebo effects, Hawthorne effects, and more generally “research participation effects” (McCambridge, Kypri and Elbourne 2014). In the other silo, while some nominally interpretivist researchers may claim that every context is unique, they routinely frame their research in reference to much broader contexts. Cohen, Manion and Morrison (2007), the writers responsible for dismissing randomised trials as ‘discredited positivism’ (above), argue that the point of case studies “is to probe deeply and to analyse intensively [...] with a view to *establishing generalizations about the wider population* to which that unit belongs” (2007:258, emphasis added).

This leads to the inevitable conclusion that aligning oneself with a particular set of assumptions, then expending energy on defending that choice to the exclusion of all others diverts energy away from sensibly addressing questions of importance. As both Ann Oakley and Ben Goldacre have emphasised, it is the research question which should dictate the approach to addressing it (Oakley 1999, Goldacre 2013), not some perceived membership of a paradigmatic tribe. By extension, it is the research question, setting, participants and expected outcomes that should inform the ontological assumptions informing the assessment of the findings of any ensuing research.

To return to the schism that was rent anew following the publication of Goldacre's provocation piece; happily we are seeing some evidence that borders have softened, and positions have mutated over time. The editors of the British Educational Research Journal, in which Elliot made the case for case studies (above), acknowledge the sea change in the conversation about how we know what we know by committing to explicitly challenging "a perception that the journal promotes any particular methodology" by keeping open debates about "the nature, scope and value of educational research" (Aldridge et al. 2018:3). ResearchED has changed its strap line from 'Working out what works' to 'Helping to bridge the gap between evidence and practice', and states that one of its key aims is "to explore 'what works' in the field of education, and to explore what the concepts contained in that statement might mean, as well as to consider the limitations of scientific enquiry in this area as well as the opportunities" (ResearchED 2018b). The EEF now emphasises the process evaluations that accompany all of its randomised trials as crucial contextual information for interpreting their findings (EEF 2018b).

This study is designed to help understand whether adopting a particular approach to teaching vocabulary to a group of children designated as EAL learners is more, less, or apparently equally effective when compared to an alternative approach. This assumes that it is possible to ascribe an objective, measurable 'effect' to a putative cause, and that the designation EAL is an objective description of a real phenomenon. Whether I believe that these statements reflect 'The Truth' or not is immaterial. Teachers operate in a world where test scores, classifications such as EAL, and assumptions of a causal relationship between what teachers do and what children learn, are all routinely accepted as 'True'. If my research is to have any relevance to teachers wishing to make informed decisions about their practice, in this instance it must also inhabit that world.

For the purposes of this study, therefore, a positivist stance and an experimental epistemology is adopted, while recognising that the findings generated by the research must be interpreted relative to the context in which it was conducted.

## 4.3 Experimental design

To address my research questions, the investigation was designed as a randomised AB/BA crossover, allocator and assessor blinded, multi-centre superiority trial. Its primary endpoints were expressive and receptive knowledge of target items of vocabulary.

### 4.3.1 Randomised crossover trial

In a randomised AB/BA crossover design (Senn 1993), the effects of alternative interventions are compared in the following way. Participants are randomly allocated to receive one of the interventions being compared, for a period of time; either intervention A or intervention B.

Outcomes are then assessed. Participants then 'cross over' and receive the intervention they did not receive in the first time period. At the end of this second period, outcomes are assessed again. This gives the design its name: some participants receive the interventions in the order AB, and other participants receive the interventions in the order BA. At the end of the trial two sets of outcomes for each participant have been generated: in the case of this study, test scores following instruction mediated by L1 and test scores following instruction in English only. These two sets of scores can then be compared using paired statistical tests of difference to assess whether a difference in average outcome between Intervention A and Intervention B is detected.

To control for possible bias associated with order effects, individual participants are randomly allocated to the order in which they receive each intervention (hence *randomised* crossover). In this study, this means that each participant was randomly allocated to receive the L1 intervention first, followed by the English-only intervention, or to receive the English-only intervention first, followed by the L1 intervention.

### 4.3.2 Rationale for the design

A randomised crossover design minimises potential threats to internal validity in the following ways. First, concealed random allocation to conditions ensures that the average of known and unknown characteristics of the participants at the time at which they are allocated to the order in which they receive the interventions differ only by chance. Second, because these unbiased groups of participants receive both interventions, but in different orders, the design minimises threats to internal validity associated with instructional order, instrument decay and maturation<sup>12</sup>. The design thus allows for more trustworthy within-participant comparisons than would be possible in a cohort

---

12. Instrument decay and maturation refers to changes that happen over time that may affect the way that participants interact with the interventions. For example, participants getting older, participants becoming more familiar with the mechanics of the intervention and the assessments, teachers becoming more familiar with the delivery of the intervention, participants becoming bored with the intervention, and so on.

pre-post design or an interrupted time series design, for example, neither of which can adequately control for this kind of bias.

Crossover designs also increase the statistical power of the study; the number of comparisons is twice that of a group randomised trial design with the same number of participants. However, if threats to validity are detected, such as significant participant dropout between phases (this can happen when participants chose to discontinue their involvement because they perceive the Phase 2 intervention to be 'worse' than the Phase 1 intervention), then a between participant comparison of the outcomes in Phase 1 can be conducted (Shadish, Cook and Campbell 2002). This approach, however, will lower the statistical power of the comparison. This is not just because the number of comparisons is fewer, but because between-participant comparisons are more prone to biases associated with unsystematic variation between participants than within-participant comparisons (Field 2018).

### 4.3.3 Limitations of crossover designs

A recognised threat to validity associated with crossover designs is the possibility of *carry-over* effects (Senn 1993). These occur when the effects of the intervention used in Phase 1 of the trial continue to exert influence during Phase 2. This compromises our ability to interpret the effects of the intervention used in the second phase. To illustrate, imagine that two approaches to reading instruction for beginner readers are being compared using a crossover design. Reading is a cumulative skill. As Castles, Rastle and Nation (2018) remind us, early in the process of learning to read, children learn that letters represent sounds (the so-called *alphabetic principle*). Later they learn the specific relationships between those letters and sounds (which letters represent which sounds), so called *phoneme-grapheme correspondence*. This is a cumulative process, to the extent that understanding the alphabetic principle is a prerequisite for understanding phoneme-grapheme correspondence. In our hypothetical crossover trial of alternative approaches to reading instruction, if the intervention used in Phase 1 successfully teaches the alphabetic principle to participants, the load on the Phase 2 intervention is lightened. It no longer has to teach the alphabetic principle, and can devote its efforts to building understanding of phoneme-grapheme correspondence. The effects of the intervention used in Phase 1 have carried over into Phase 2, making it impossible to disentangle the differential effects of the two interventions. The outcome assessment administered at the end of the second phase will not be a measure of the effectiveness of the intervention used in Phase 2, but a measure of the effects of both interventions combined.

This makes crossover trials unsuitable for comparing educational interventions that build competence through accumulation of interdependent skills. There appears to be no *prima facie*

reason to assume that learning independent items of vocabulary is a cumulative skill, and therefore no reason to assume that there would be carry over effects from Phase 1 to Phase 2. Therefore, I did not consider carry-over to be a potential threat in this comparison.

That said, because I could not completely rule out the possibility that understanding of a word learned in Phase 1 might affect the learning of a word taught in Phase 2 (for example if words are cognates, like *settler* and *settlement*), I addressed this by teaching the target words in different random orders at each school. More discussion of possible effects of instructional order and the measures I took to account for bias associated with them are detailed in Section 4.6.4.

## 4.4. Setting of the study, schools and participants

The study was conducted with children in Years 4, 5 and 6 (aged 8-11), classified as learning EAL, at four state-maintained primary schools in the city of Oxford, UK.

### 4.4.1 The City of Oxford

Oxford is a city of approximately 160,000 residents (Oxford City Council 2017) in the South East of England with greater ethnic and linguistic diversity than is the average for England as a whole (Office for National Statistics 2013). The 2011 Census reported that 22% of Oxford residents were from a black or ethnic minority group, with a further 14% having a white, non-British ethnic background (Oxford City Council 2016a). Migration has made an important contribution to Oxford's diverse character, with 30% of the city's population having been born outside the UK (Oxford City Council 2017). The 2011 census reports that approximately one quarter of all residents born outside the UK arrived in Oxford when they were under the age of 16, and therefore still of compulsory school age. The City Council also reports that children born in Oxford to mothers who were born overseas has increased from 19% in 2001 to more than 50% in 2016 (Oxford City Council 2016b, 2017). This suggests that, for this considerable segment of Oxford's school population, family language practices are likely to include languages other than English.

In addition to relatively recently arrived migrants the city has a large, well established, Asian community. In the 2011 census 12% of the city's population were classified as having either Indian, Pakistani, Bangladeshi, Chinese or 'Other Asian' backgrounds (Oxford City Council 2016a).

The two universities in the city attract a large number of scholars and academics from outside the UK. At the time of writing, 41% of the students at the University of Oxford and 48% of its academic staff are from countries outside of the UK, representing 138 different nationalities (University of Oxford 2016a, 2016b). Oxford Brookes University receives in excess of 3000 students from more than

130 different countries every year, representing 18% of its student population (Oxford Brookes University 2016a, 2016b). Many of these students and academics come to live in the city with their families and send their children to local schools.

In the 2011 census, 16.2% of respondents in Oxford indicated that their main language was not English, twice the national average (Office for National Statistics, no date, Oxford City Council 2016c). The equivalent figure for children attending the city's state maintained schools is considerably larger at 27% (Oxford City Council 2013). These data, collected in 2010, state that after English the most commonly spoken home languages by the city's school children are Panjabi, Urdu, Bengali, Arabic and Portuguese, and that, in total, approximately 120 different languages are represented. Given national trends over the intervening seven years (see NALDIC 2016) we might expect those numbers to have increased dramatically since the last census. Oxford's linguistic diversity thus provided an ideal setting to investigate the research questions for this project.

## 4.4.2 The Schools

### 4.4.2.1 Sampling frame and sampling technique

The population for the study was upper-primary-school-children (children in Years 4 to 6, aged 8 to 11) classified as having EAL, in state-maintained primary schools in England. Random sampling of this population, while methodologically preferable, would be impossible due to financial, logistic and practical constraints. Instead, a convenience sample was taken, based on the linguistic profile, availability, and willingness to participate of local schools.

### 4.4.2.2 Finding eligible schools

Three key characteristics informed how the schools invited to participate were selected. Firstly, schools needed to have sufficiently large numbers of pupils for whom English is an additional language to allow for a meaningful comparison to be made while keeping the logistics of visiting multiple sites manageable. Second, schools needed to be located within the city limits to allow me, the principal investigator and individual responsible for delivering the intervention, to visit them easily. Thirdly, based on the time available to conduct the investigation, the number of schools I was able to work with would necessarily be limited.

There are 44 state maintained primary schools that come under the jurisdiction of the City of Oxford (Oxfordshire County Council 2016). Twenty-five of these are located inside the city's encircling ring road, which, for the purposes of this project, I have adopted as a proxy for the city limits proper. Precise data about the numbers of children classified as EAL in the city's schools is not readily

available, so to assess how likely these schools were to have a sufficiently large number of eligible children, I referred to their most recent Ofsted<sup>13</sup> reports.

Ofsted reports do not report the numbers for EAL learners in individual schools. However, they often include comments indicating the extent to which their EAL population relates to the national average - which is 16.2% (Strand et al. 2015). This included phrasing such as “the number of pupils who speak English as an additional language is higher than the national average” (Ofsted 2012a:4), as well as more precise language, such as “70% of pupils speak English as an additional language” (Ofsted 2015:3). Conversely, some reports used wording that suggested that a school would be unlikely to meet my criterion of above average proportions of EAL learners. For example, one school was described as being “well below the national average” (Ofsted 2014:3), and of one school Ofsted said only that “[a] large majority of pupils are of White British heritage.” (Ofsted 2012b:3). I read the Ofsted reports for all schools located within the ring road and excluded any that indicated that the proportion of EAL learners was below the national average. In addition, I excluded schools for whom Ofsted had used terms like ‘just’ and ‘slightly’ above the national average. This left 23 schools described in terms that could be unequivocally interpreted to mean that the proportion of children with EAL was above the national average. These schools constituted a longlist of potential participants.

#### 4.4.2.3 Sample size

In a trial with sufficient resources to allow for a functionally unlimited number of participants, a power calculation based on the hypothesised degree of difference between scores on the outcome measures that would be needed to assert any differential effect of the intervention would have been conducted. This would then have guided the estimation of the ideal number of participants needed and therefore the way by which schools were recruited. A correspondingly sized random sample of schools meeting the language profile requirement would then have been taken. In the context of this study, which has no budget and is reliant on a single individual to deliver the alternative interventions, a different approach was necessary.

The paucity of similar research upon which to inform estimates of the sample size needed to detect an effect of a declared magnitude, should one exist, and the relatively subjective nature of what the size and direction of an effect needs to be for that effect to be considered educationally meaningful, complicates attempts at power calculations. That said, one of the studies that informed the design of

---

13. Ofsted, the Office for Standards in Education, Children’s Services and Skills inspects and regulates all state maintained schools in the country. It publishes reports of the findings of its inspections (Gov.UK, no date).

this study (Lugo-Neris et al. 2010) detected a statistically significant effect in a sample of just 22 children. A similar study, also included in my review, with Korean primary school children (Lee and Macaro 2013) included approximately<sup>14</sup> 15 cases (units of allocation in this non-randomized comparison were school classes, not individual pupils). A statistically significant difference in outcome between groups was detected in this study also. Based on these observations, and the rule of thumb that larger is better (Field 2018), I aimed to recruit no fewer than 22 participants, and ideally as many as possible.

Based on my review of the Ofsted reports for primary schools in Oxford, and the likely number of EAL learners in each, I estimated that a minimum of four schools would be required to generate sufficient numbers of participants, assuming that not all of the EAL learners in those schools would agree to participate. Because I also assumed that not all invited schools would agree to participate, I created a shortlist of eight from the pool of 23 suitable schools, and invited them to participate. In addition to information about the proportion of EAL learners in these schools, the shortlist was informed by professional relationships I enjoy with teachers and governors in schools in Oxford, developed through my role as convener of a Professional Network of EAL teachers in Oxfordshire. I was personally acquainted with staff members in six of the eight schools on the shortlist, and considered these likely to be well disposed to participation. While the sampling frame was all children in state-maintained primary schools in the City of Oxford classified as EAL, the sampling technique is, therefore, best described as a convenience sample.

#### 4.4.2.4 Recruiting schools

In November 2016, I sent a letter of invitation and outline of the project to the headteacher of each shortlisted school. One headteacher responded shortly afterwards to express interest in participating. The remaining seven headteachers did not respond at all. At the start of the new term in January 2017 I emailed the headteachers who had not responded to reiterate my invitation. A representative of one school responded at this point to say that, in principle, the school was happy to allow me to collect data there but that it was not prepared to offer a link member of the school staff with whom I could liaise. This condition would have made conducting the study almost impossible. I thanked them for their offer but did not recruit them. Three schools did not respond at all, so no further attempts to recruit them were made. The headteachers of the remaining three

---

14. It is unclear from Lee and Macaro's report how many cases were involved in the comparison, as they report only the total number of participants (n=443). By dividing 443 by 28, the average class size in Korean primary schools at the time of their study (OECD 2012), I derived an estimate of the number of cases in the study.



schools responded to express interest in participation. The four participating schools will be referred to as School 1, School 2, School 3, and School 4 throughout this document.

Having secured initial expressions of interest from the headteachers of four schools, I visited each and gave a presentation to (variously) headteachers, EAL coordinators, EAL support teachers, Deputy and Assistant Headteachers, Special Educational Needs coordinators and class teachers to describe the project and outline what participation would entail. Representatives in all schools were enthusiastic about their involvement with this study specifically, and about involvement in research that addresses the needs of EAL learners in general. At this time, a 'link teacher' was established at each school as a point of contact and to help with administration of the trial. The link teachers in School 1 and School 2 were the headteachers; in School 3 the link teacher was the school's Special Needs Coordinator; and in School 4 the link teacher was the school's EAL Specialist. In the four schools that agreed to participate, a total of 208 children in Years 4, 5 and 6, representing 33 different L1s were classified as EAL and were thus eligible to participate.

#### 4.4.2.5 Characteristics of participating schools

Unless otherwise stated, information in this section was obtained directly from the schools or from their Ofsted reports. To preserve the anonymity of the schools, the Ofsted reports for these schools have not been included in the references.

##### **School 1**

School 1 was a voluntary controlled Church of England primary school with 315 pupils on roll. The majority of the school's catchment area is listed on the Income Deprivation Affecting Children Index (IDACI) (MHCLG 2015) as amongst the 40% most deprived neighbourhoods in the country. A small proportion of the catchment falls within a slightly less deprived area (amongst the 50% most deprived). The proportion of pupils eligible for the pupil premium<sup>15</sup> is above the national average. The school meets the minimum expectations set by government for pupil attainment and progress. It was rated 'Good' by Ofsted at its last inspection. It is a culturally and linguistically diverse school, with about two thirds of pupils classified as belonging to a minority ethnic group. The largest of these groups is children from Pakistani backgrounds, though children represent a wide variety of cultural and linguistic heritages. At the time of the study 16 different languages were spoken among eligible pupils, including languages from Asia, Africa, Europe and South America. Approximately 45%

---

15. The pupil premium is additional funding made available to publicly funded schools to help raise the attainment of disadvantaged pupils. Pupil premium funding is awarded based on the number of children at the school who have been registered as eligible for free school meals at any point in the previous six years. (DfE 2017b).

of pupils were classified as EAL. One of the Early Years teachers serves as EAL coordinator. No special arrangements appear to be in place for teaching children with EAL. Linguistic and cultural diversity at the school is celebrated through multilingual signage and a map of the world in the foyer, displaying the countries of origin represented among pupils.

### **School 2**

School 2 is a Voluntary Aided Church of England primary school with 349 pupils on roll. The school's catchment area extends over four neighbourhoods, each with a different ranking on the IDACI. Two are ranked as among the 30% and 40% most deprived areas in England, respectively. The other two neighbourhoods are ranked as among the 30% and 20% least deprived areas in England, respectively. The proportion of pupils eligible for the pupil premium is below the national average. The school meets minimum expectations set by government for pupils' attainment and progress. It was rated 'Good' by Ofsted at its last inspection. About half of pupils are from minority ethnic backgrounds. Twenty-one different L1s from four continents were represented by eligible participants. About 30% of pupils were classified as EAL. The school has an EAL coordinator who is also a class teacher. No special arrangements appear to be in place for teaching children with EAL. Cultural and linguistic diversity is acknowledged by a sign in the foyer that says 'Welcome' in a variety of different languages and celebrations of the cultural diversity of the school, such as a pupil-led Bollywood Dancing show that was being advertised while I was conducting the study.

### **School 3**

School 3 is a Voluntary Aided Church of England primary school with 417 pupils on roll. The school's catchment extends over four neighbourhoods, two of which are ranked on the IDACI as amongst the 10% least deprived areas in England, and two of which are ranked as amongst the 20% least deprived areas. The number of pupils supported by pupil premium funding is below the national average. Ofsted reports that all pupils (including those with special education needs, EAL, and pupils known to be eligible for free school meals) achieve well at the school. It was rated 'Good' by Ofsted at its last inspection. The proportion of pupils from minority ethnic backgrounds is above the national average, but the majority are White British, White Other, or Mixed-Heritage. Mobility in the school is comparatively high. Many students transfer to the independent sector during the school year, and there are a number of 'sojourning' pupils - those attending the school while their parents study or work temporarily in Oxford. Fifteen different L1s were represented by eligible participants. The majority of these were European languages, with a small minority of speakers of Hebrew, Korean and Mandarin Chinese. The number of children classified as EAL was approximately 26%.

Responsibility for EAL comes under the jurisdiction of the Special Educational Needs coordinator. There was little evidence around the school to suggest that explicit support for EAL learners or promotion of multilingualism is part of the daily business of the school.

#### **School 4**

School 4 is a non-denominational community primary school with 319 pupils on roll. The school's catchment area extends over six neighbourhoods. The three closest neighbourhoods are ranked amongst the 30% and 40% most deprived areas in England on the IDACI, while neighbourhoods on the periphery of the catchment are among the 50%, 40% and 30% least deprived. Pupil premium funding is above the national average. The school was rated 'Good' at its last Ofsted inspection, the report stating that, from low starting points, most pupils reach average attainment in reading, writing and maths by Year 6. Eighty percent of the pupils come from a wide variety of minority ethnic backgrounds. The largest proportion (about a quarter) are of Pakistani heritage. Thirty different L1s were represented at the school, the majority of these being South Asian languages (Urdu, Panjabi, Bengali, Pashto), with smaller numbers of speakers of languages from Europe, Africa, the Middle East, East Asia, and South America. The number of pupils joining the school part-way through their primary education is higher than the national average. Ofsted reports that some of these children have had no prior experience of formal education, and many arrive with little or no English. Approximately 65% of pupils were classified as EAL. The EAL specialist teacher at the school was the only staff member of all participating schools with formal qualifications for teaching EAL learners, having undertaken comparatively extensive professional development and training during previous employment in London. Her sole responsibility is support for EAL learners in the school across all year groups. This is done through in-class support and withdrawal sessions for the least proficient users of English. She actively promotes multilingualism, most visibly through a bilingual library that she wheels into the playground once a week to allow parents (who are often reluctant to come into the school building) to choose books to read with their children in their L1s.

### **4.4.3 Participants**

#### **4.4.3.1 Recruiting participants**

Initial eligibility criteria for participation in the study were that participants must be in Years 4, 5 or 6 (aged between 8 and 11) and be classified in their school records as having EAL, and be at least minimally proficient in their non-English language. More information about assessing minimal proficiency in pupils' non-English languages was presented in section 4.4.3.3.

Once approval from the headteachers had been obtained, I invited eligible pupils and their parents to a meeting held at each school where I described the project, the rationale and existing evidence-base that informed its conception, and what participation would entail. I prepared a participant information sheets for both parents and children (Appendices C and D) and consent form (Appendix E), which I distributed to attendees of the meetings.

Turnout at the meetings was low relative to the number of eligible pupils at each school. Attendance ranged from representatives of 10 families in the meeting held at School 4, to zero attendees at the meeting held at School 3. Parents of eligible children who did not attend the meeting were sent a copy of the participant information sheet and a consent form. To allow ample time to consider and respond to the invitation, parents were asked to return consent forms within a week of receiving the original invitation.

Following the meetings, I asked class teachers to remind eligible students of the invitation to participate and to encourage them to talk to their parents and return the consent forms if they wished to take part. School 4 reported that parents of EAL learners were often reluctant to come into the school building, and that responses to written communication from the school were not always forthcoming. Therefore, the link teacher at School 4 assisted in the recruitment process by meeting parents in the playground as they dropped off or picked up their children. She told them about the project and invited them to consent to their child's participation. She also made and displayed on the school notice board a poster advertising the project. School 2 followed up initial invitations to participate by sending a reminder email to the parents of eligible students using the school's ParentMail system. No modifications to the recruitment process were made by Schools 1 and 3.

At the end of the recruitment window, the parents of 47 pupils had given their consent, 23% of all eligible pupils at the participating schools. Prior to randomisation, seven of these students were excluded from the study. One student had been recruited in error from the wrong year group, one was withdrawn at the request of the school, one left the school soon after recruitment, one was unable to participate due to timetable clashes with the intervention sessions, and three were excluded because no materials in their L1s were available. There were no more dropouts, leaving 40 students who participated in full. See Figure 4.1 for the flow of participants through the trial.

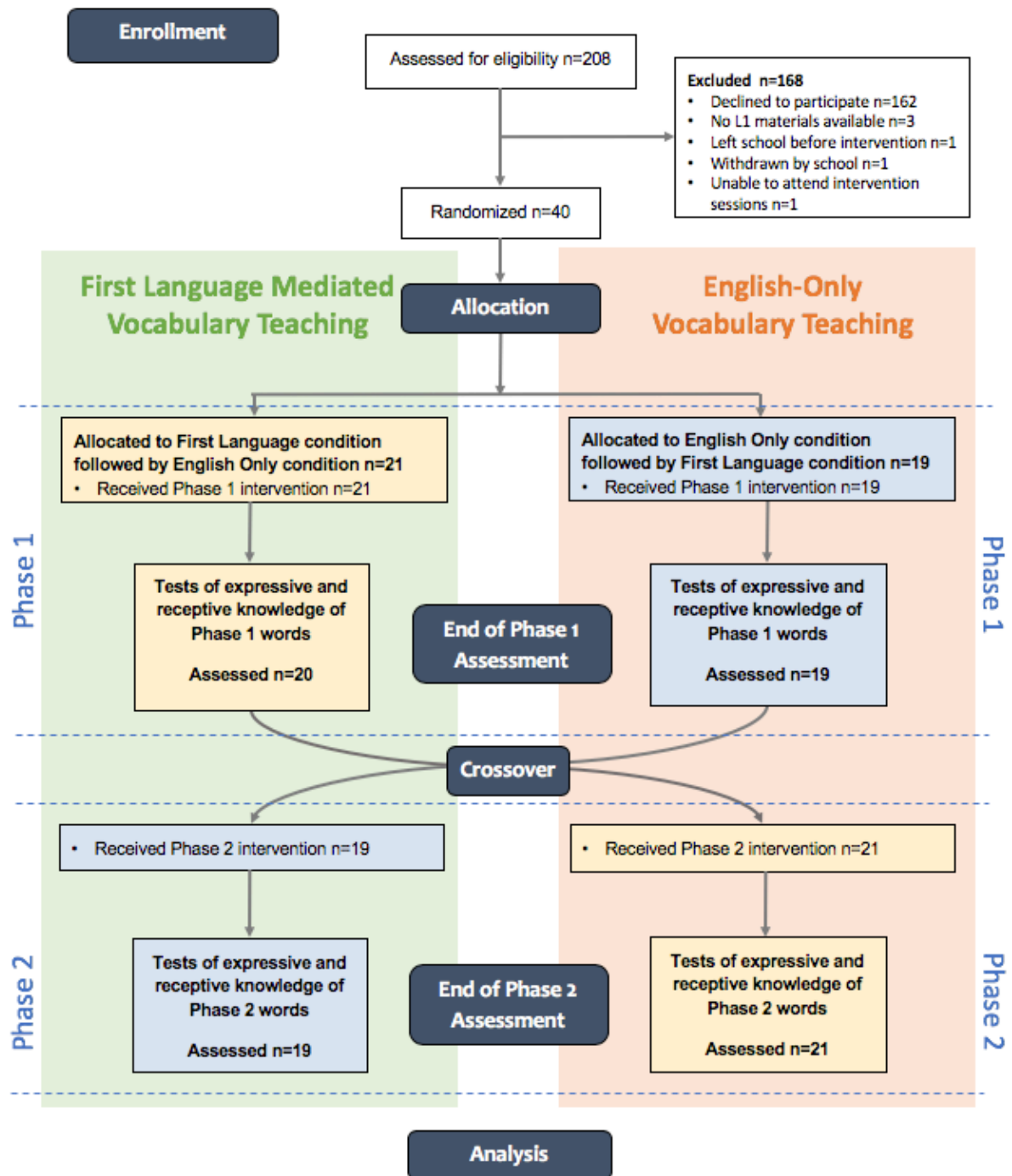


Figure 4.1 CONSORT<sup>16</sup> Flow diagram.

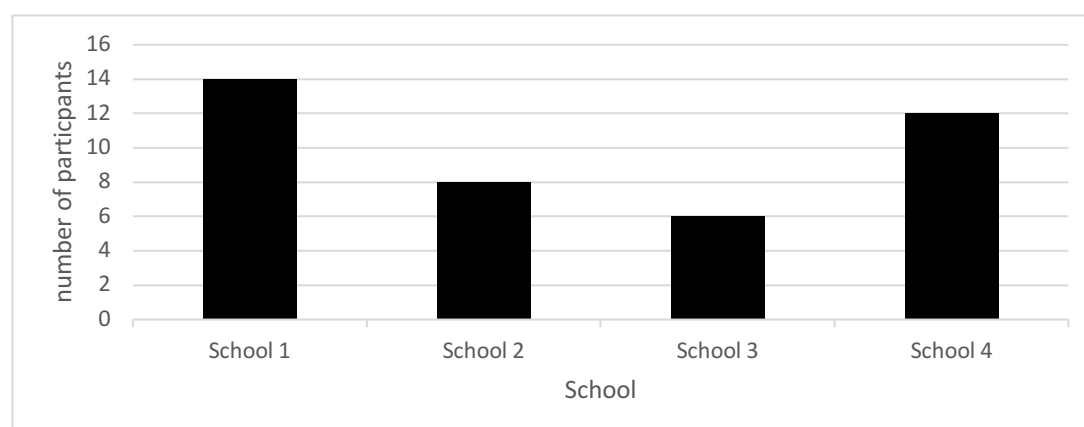
<sup>16</sup> CONSORT stands for Consolidated Standards of Reporting Trials. The CONSORT flow diagram is part of a suite of guidance materials designed to offer “a standard way for authors to prepare reports of trial findings, facilitating their complete and transparent reporting, and aiding their critical appraisal and interpretation.” (CONSORT no date).

#### 4.4.3.2 Participants

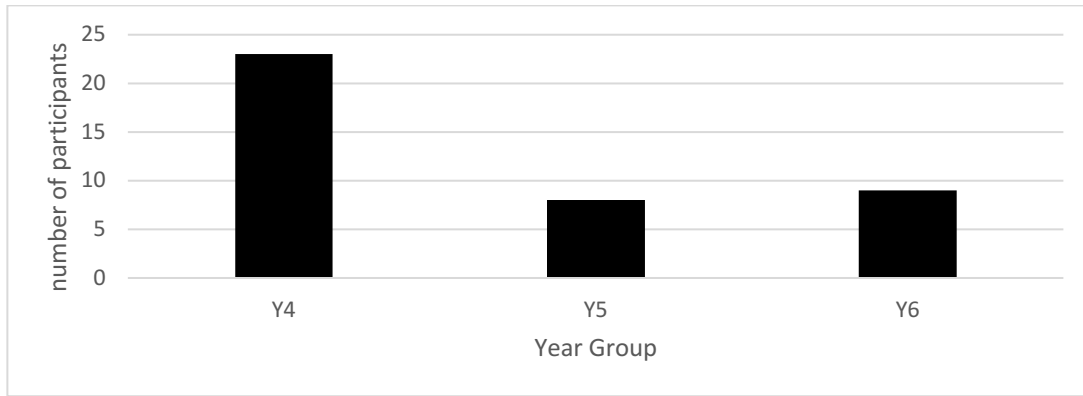
Forty children in Years 4, 5 and 6 (ages 8 to 11 years), classified by their schools as learning English as an additional language (EAL) in four state-maintained primary schools in the City of Oxford, took part in the trial.

I collected descriptive information about the participants before initiating the intervention. Some of this information was obtained through direct observation (*e.g.* year group and gender), some was obtained from school records (*e.g.* English proficiency, stated L1 of the child), and some was collected directly from the participants using a questionnaire (*e.g.* educational experience).

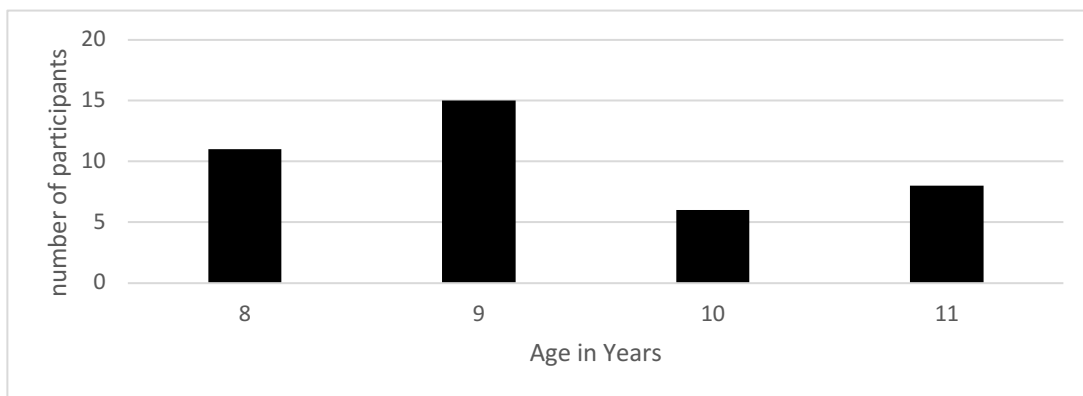
There were fourteen participants in School 1, eight in School 2, six in School 3, and twelve in School 4 (Figure 4.2). More than half of all participants were in Year 4 ( $n=23$ ), with eight participants in Year 5 and nine participants in Year 6 (Figure 4.3.) Median age was 9 years (Figure 4.4). Twenty-four participants were girls and 16 were boys. Between them, participants spoke 14 different L1s (Figure 4.5).



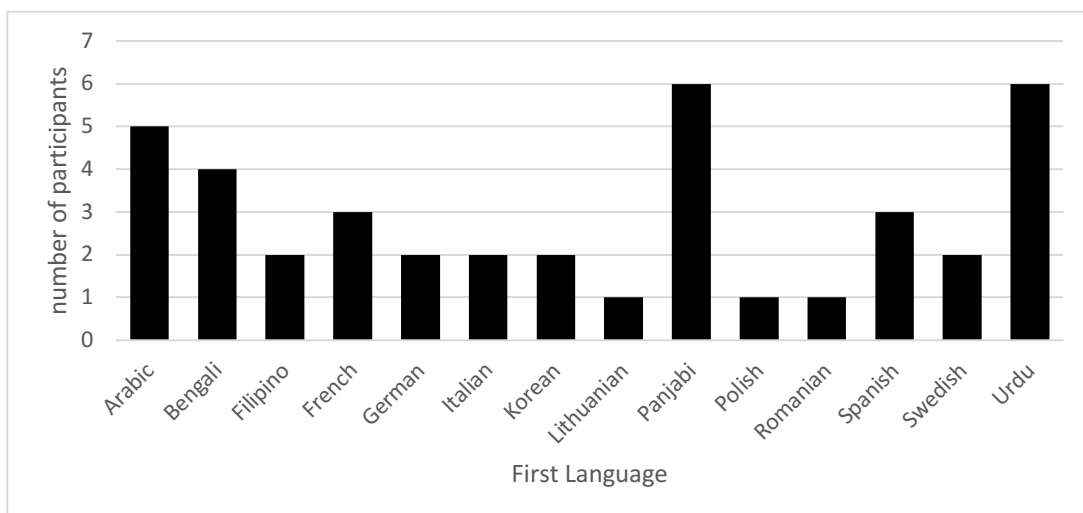
**Figure 4.2.** Number of participants by school.



**Figure 4.3.** Number of participants by year group.



**Figure 4.4.** Number of participants by age in years.



**Figure 4.5.** Number of participants by L1.

#### 4.4.3.3 Language proficiency of the participants

The classification EAL is very broad. The most recent DfE definition was published in the 2014 DfE guidance to completing the school census. It states that:

“A first language other than English should be recorded where a child was exposed to the language during early development and continues to be exposed to this language in the home or in the community [...] If a child was exposed to more than one language (which may include English) during early development the language other than English should be recorded, irrespective of the child's proficiency in English. In the case of an older pupil who is no longer exposed to the first language in the home, and who now uses only another language, the school should consult with the pupil or parent to determine which language should be recorded.” (DfE 2014:38).

This means that the classification includes everything from children who are equally (and possibly highly) proficient in two or more languages, through children who are proficient in their L1 and who are still learning English, and children who, while exposed to a language other than English in infancy, no longer use or understand it and who, therefore, might be functionally indistinguishable from monolingual English-speaking children. Clearly, used in isolation, the term EAL is not a helpful indicator of whether a child might benefit from this intervention. For example, a sample of EAL learners may include children who do not understand their ‘L1’ and who presumably, therefore, cannot hope to gain much from pedagogical input in that language.

To help ensure that children who appeared eligible on the basis of school records were capable of engaging with the L1 intervention materials, a sense of their L1 proficiency was needed. Ideally, each potentially eligible participant in this study would have been assessed for L1 proficiency using a formal, standardised test. This would have provided an estimate of how likely they would be to understand the L1 materials used in the comparison, and thus an indication of whether the intervention might help them. However, the initial pool of potentially eligible participants consisted of 208 children representing at least 39 different languages<sup>17</sup>. It was beyond the capacity of this PhD project to source and acquire appropriate and mutually compatible L1 assessments or the personnel needed to conduct them in all of the languages represented. As a proxy for this ideal, participants completed a language background questionnaire (Appendix F).

---

17. I say ‘at least’ here because a number of the languages included in the school records were not individual languages, but language groups. For example there were children classified as speaking Chinese, but no indication of whether this was Mandarin, Cantonese, Hokien, Haka, or any other of the 275 indigenous living languages of China (Simons and Fennig 2017).



#### 4.4.3.4 Procedure for carrying out the language background questionnaire

The questionnaire was created using Google Forms and was conducted online at the participating schools. Participants were asked to judge their proficiency in the four domains of speaking, listening, writing and reading in their L1. They rated these on a five-point Likert scale. In addition, participants were asked for information about their educational experiences in both L1 and English. These were: length of residence in England, completed years of schooling in England, whether they took L1 literacy lessons outside school, and whether they had ever been to a school in which their L1 was used as the principal medium of instruction. A multiple-choice format was used for these questions, with spaces for additional comments if required.

I supervised each questionnaire session. Participants were provided with a desktop pc, a laptop, or a tablet computer (depending on the resources available at the school). They accessed the questionnaire online using laptop computers or iPads and were told that there was no time limit on completing it. I explained the purpose of the questionnaire and demonstrated the process of answering each question type. Similar information and instructions were presented in writing on the questionnaire. I asked participants to answer as many questions as they could, but told them that they were free not to answer any question if they did not want to. I read the questions to any children who asked me to, and I discussed their response options with them if they wanted me to. The nature of these interactions included, for example, children seeking to clarify whether being able to recite passages from the Koran constituted being able to read in Arabic, or whether watching TV in their L1 was the same as understanding it.

A common request for clarification concerned a phrase used on two of the Likert scales: 'I can speak/understand it perfectly'. I clarified this to mean 'I can say/understand anything that I *want* or *need* to say or understand'. The nature of language proficiency (or 'fluency' as it is often incorrectly termed, or 'perfection' as I chose to style it for the purpose of this questionnaire) is not straightforward. After all, I consider myself to be a highly proficient user of English, but would be hard pressed to maintain that opinion in a discussion about quantum mechanics with a group of physics majors. The context is, as in so many cases, all important. Valdés and Figueroa (1994:34) explain that "Knowing a language and knowing how to use a language involves a mastery and control of a large number of interdependent components and elements that interact with one another and that are *affected by the nature of the situation in which communication takes place*" (emphasis added). For the purpose of this study, therefore, the participants' assessment of their L1 proficiency was informed by the context in which they usually used it. As with all self-reported data, these kinds of judgement are inevitably subjective and therefore not necessarily externally valid. Self-reported

language proficiency based on the definition given above is a good illustration of the threats to validity associated with this method of data collection. Nonetheless, in the absence of an objective measure of L1 proficiency it was the best available alternative.

The responses to the language background questionnaire were then taken into account when determining whether a child met the minimal L1 proficiency requirement for participation. Children who indicated that they were not at all proficient in their L1 (and specifically in listening, as the L1 materials were audio recordings) would have been excluded from the study. However, in the event, all children indicated at least minimal proficiency, and an overwhelming majority indicated high levels of speaking and understanding proficiency. No children were excluded on the basis of these factors.

#### 4.4.3.5 Summary of the information collected using the language background questionnaire

All participants reported being at least minimally proficient in listening in their L1 (no participants reported being unable to understand their L1 at all). A large majority (n=31, 77.5%) reported that they spoke and understood their L1 'well' or 'very well'. There was a greater spread of proficiency in reading and writing, with just under a quarter of participants reporting that they could not read or write their L1 at all (Table 4.1).

**Table 4.1** Self-reported level of L1 proficiency in the four language domains of Speaking, Listening, Writing and Reading.

	1 (Not At All)	2	3	4	5 (Very Well)
How well do you SPEAK your home language? n(%)	1 (2.5%)	1 (2.5%)	7 (17.5%)	16 (40%)	15 (37.5%)
How well do you understand your home language when you HEAR it? n(%)	0 (0%)	2 (5%)	2 (5%)	17 (42.5%)	19 (47.5%)
How good are you at WRITING in your home language? n(%)	9 (22.5%)	8 (20%)	9 (22.5%)	8 (20%)	6 (15%)
How good are you at READING in your home language? n(%)	7 (17.5%)	3 (7.5%)	11 (27.5%)	7 (17.5%)	12 (30%)

#### 4.4.3.6 Linguistic and educational background

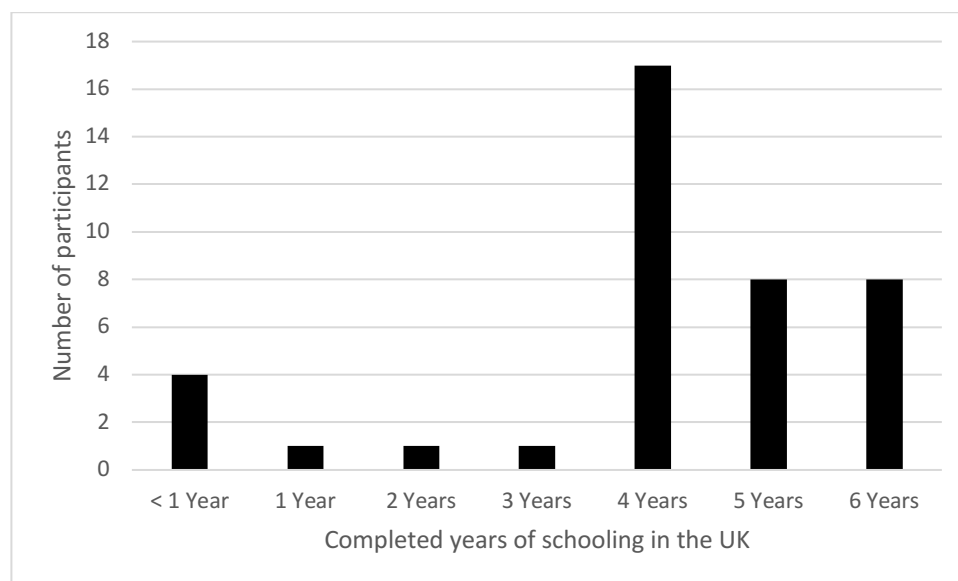
Participants were asked about the length of time they had lived in England, the amount of time they had been in the English school system, and whether they had ever been to school in a country where their L1 is the main language of education. These data were gathered to address Research Question 2, to determine whether these pupil characteristics were differentially associated with the outcomes of the interventions being compared. For example, a student who had received some of their education in schools in which their L1 was routinely used might be in a better position to draw on their existing academic schemata when primed to do so using their L1. Conversely, EAL learners who had received the majority of their academic instruction in English may not benefit so much from input of this type.

Most participants (n=27, 77.5%) were born in England. Four participants had arrived in England before they reached school age, and two had arrived at the time their formal schooling began in Reception Year. The majority (n=33, 82.5%), therefore, had received all of their formal schooling in the English school system (Table 4.2). The remaining 7 participants arrived in England at various

points in their educational career, and thus had experienced at least some of their formal education in their home country. However, two of these participants had attended schools in their home countries in which the language of instruction was English, leaving 5 participants for whom some formal experience with academic language in their L1 can be confidently assumed. Four participants had arrived in England at the beginning of the school year in which the intervention was conducted, and thus had been in the English system for less than a year (Table 4.2 and Figure 4.6).

**Table 4.2** School year on arrival in England

School Year on Arrival	Born in England	Before school age	Yr R	Yr 1	Yr 2	Yr 3	Yr 4	Yr 5	Yr 6
Number of pupils	27	4	2	1	1	1	3	0	1



**Figure 4.6** Participants' completed years in the English education system.

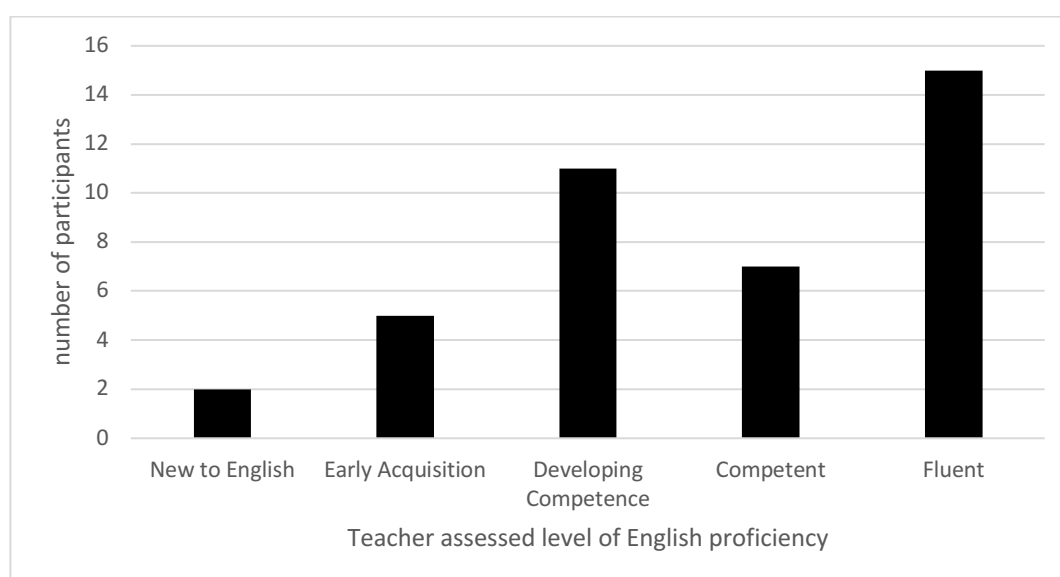
Fourteen participants (35%) reported having special L1 literacy lessons outside of school. The nature of these lessons varied, including informal lessons with a family member or family friend, lessons from a teacher who came to their house, and attendance at a community or complementary school.

#### 4.4.3.7 English language proficiency

In addition to obtaining a gauge of participants' proficiency in their respective L1s, I also collected information about participants' English language proficiency. This was taken from the schools' routine assessment data.

In September 2016 the UK's Department for Education (DfE) introduced the requirement for schools to report levels of English language proficiency for each child classified as EAL (DfE 2016a). A five-point scale was introduced, against which teachers were asked to indicate the English proficiency of their EAL pupils. The scale consists of five broadly defined stages: New to English, Early Acquisition, Developing Competence, Competent, and Fluent. These were adapted from a similar scale used in Wales, and will feel familiar to anyone with experience of equivalent tools used in North America, such as SOLOM (Student Oral Language Observation Matrix) (see Dennis et al. 2018). Here, broadly the same five stages are described as: Pre-Production, Early Production, Speech Emergence, Intermediate Fluency, and Advanced Fluency (after Krashen and Terrell 1983).

These data were obtained directly from the school records (see Figure 4.7).



**Figure 4.7.** Participants' English language proficiency levels.

The information about English proficiency was not used to inform decisions about the eligibility of any child to participate. As noted in my discussion of the literature, use of L1 (however defined) is assumed by many to be good practice at all levels of bilingualism (e.g. García and Sánchez 2015). On that basis, no child was considered, *a priori*, unsuitable for participation because they had been

assessed by their teachers as having high levels of English language proficiency. Rather, this information was collected for regression analysis of possible interactions between levels of English proficiency and scores on the study's primary outcome measures. This would help to determine whether L1 as used in the intervention was differentially effective for children at different stages of acquiring English.

## 4.5 Allocation to comparison conditions

Participants were stratified by school, then allocated to the order in which they received the two interventions, using a well-concealed, unbiased allocation schedule. Allocation was at the level of the individual. The relatively small number of participants, and the skew towards Year 4 children and children with higher levels of English proficiency suggested that further stratification would add little to the generation of unbiased comparison groups.

A third party, familiar with the nature of the project, but with no knowledge of the schools or the children recruited to the trial, performed the allocation. He was given four sets of numbers, each consisting of the Unique Identifying Numbers (UINs) of the participants at each of the four schools. Names of the participants were not used, to ensure that the allocator was unable to guess information about them such as their ethnicity, gender, and L1. This protected against conscious or unconscious bias in, and subversion of, the allocation process. Taking one set of numbers for each school at a time, the allocator entered the UINs into the random sequence generator on the website [random.org](http://random.org)<sup>18</sup> (Random.org 1998–2017) and set it to randomly sort the UINs into two lists. The first list constituted the children in that school who would receive the English-only condition in Phase 1 of the study and who would subsequently cross over to the L1 condition in Phase 2. The second list constituted the children in that school who would receive the L1 condition in Phase 1 of the study and who would subsequently cross over to the English-only condition in Phase 2. Characteristics following randomization are presented in Table 4.3. The process was repeated for each school. Group allocation was concealed from participants and their class teachers until the first intervention session.

---

18. Random.org generates *true* random numbers via atmospheric noise rather than the pseudo-random numbers generated by software such as Excel.

**Table 4.3** Participant characteristics after random allocation to intervention order

		Intervention Order	
		L1, English-only	English-only, L1
<b>Number of participants</b>		19	21
<b>Year Group</b>	Year 4	14	9
	Year 5	5	3
	Year 6	2	7
<b>Age (years)</b>	8	7	4
	9	9	6
	10	4	2
	11	1	7
<b>English Level (teacher assessed)</b>	New to English	0	2
	Early Acquisition	3	2
	Developing		
	Competence	8	3
	Competent	2	5
	Fluent	8	7
<b>Mean L1 Proficiency Rating</b>		3.7	3.6
<b>Gender</b>	Male	9	7
	Female	12	12

## 4.6 Interventions

### 4.6.1 Overview

The trial was intended to assess whether students' expressive and receptive knowledge of a set of academic vocabulary items drawn from the English national curriculum was affected by providing pedagogical input related to those words in their L1s, compared with providing that input using only English.

The design of the interventions was informed by the procedures used in four of the studies included in my systematic review (Sieh 2008, Lugo-Neris et al. 2010, Lee and Macaro 2013 and Camo and

Ballester 215). In these studies, the effects of providing information about target items of English vocabulary, either in the L1 of the participants or in English, was assessed. In those studies, however, all participants shared an L1 with their teachers. Teachers in these studies, therefore, were able to provide L1 input directly and efficiently to all of their students simultaneously. In my study, 14 different L1s were represented. Like many teachers in England, my linguistic repertoire is limited to many fewer than 14 languages. As the individual responsible for delivering the interventions, I was unable to mirror the approach taken in those previous studies. One possible way around this would have been to work with teaching assistants proficient in the languages represented by my participants to deliver the intervention. However, this was neither practically or financially feasible, nor reflective of the capabilities of typical English primary schools, whose practice I hope to inform. My intervention, therefore, required an alternative method by which to provide the appropriate input. To do this I prepared videos in which information about the target items of vocabulary was provided either in English or in the L1 of each participant.

Information about how the videos were made and the choice of target vocabulary is detailed in Section 6.3. The procedures adopted for the trial are detailed in the following section.

#### 4.6.2 Procedure

Intervention sessions were conducted in normal school hours, and participants were withdrawn from their usual classes in order to take part. Each session lasted for one hour and focused on three items of vocabulary. Individually, participants used a laptop or tablet computer and headphones to view a short video that explained the meaning of an item of vocabulary that they were told was important to know for their learning at school (the target word). The video included information about characteristics of the target word, examples of the target word, non-examples of the target word, and a succinct definition of the word. This information represents the key understanding needed to demonstrate concept mastery according to Frayer et al (1969), which I will discuss in more detail in Section 4.6.3.1. (below). Participants had control of the video and were able to stop, start, rewind and review them as they wished. All videos can be viewed online at <https://vlp2017.wordpress.com/>.

After watching each video, participants engaged in a teacher-led discussion, in English, about the target word. This discussion reiterated the four key areas of concept mastery. Participants were asked to recall what they had learned from the video, describe the characteristics of the target word, add further examples of the target word to the ones given in the video, suggest additional non-examples of the target word, and orally rehearse their own definition of it. Participants then completed a concept map designed to help them structure and organise this understanding of the



word. Using the concept map, participants wrote down characteristics, examples, non-examples and a definition of the word. This process was repeated for each of the three words taught in each session.

The interventions differed between L1 and English-only conditions in two ways. First, in the L1 condition the voiceover of the videos was in the relevant L1 of the participant, while in the English-only condition it was in English. Second, in the L1 condition the participants were given the option to complete the concept map using their L1 or a combination of L1 and English. In the English-only condition, they were instructed to use only English. Otherwise the interventions were identical.

Each phase of the study lasted two weeks. In each two-week period participants attended three separate teaching sessions until they had studied nine target words. On the fourth session, participants' expressive and receptive understanding of the nine words was assessed using gap-fill and multiple-choice format quizzes, conducted in English (see Appendices G and H). Participants then crossed over to the second phase to repeat the entire process with the comparison intervention and a new set of nine words.

### 4.6.3 Materials

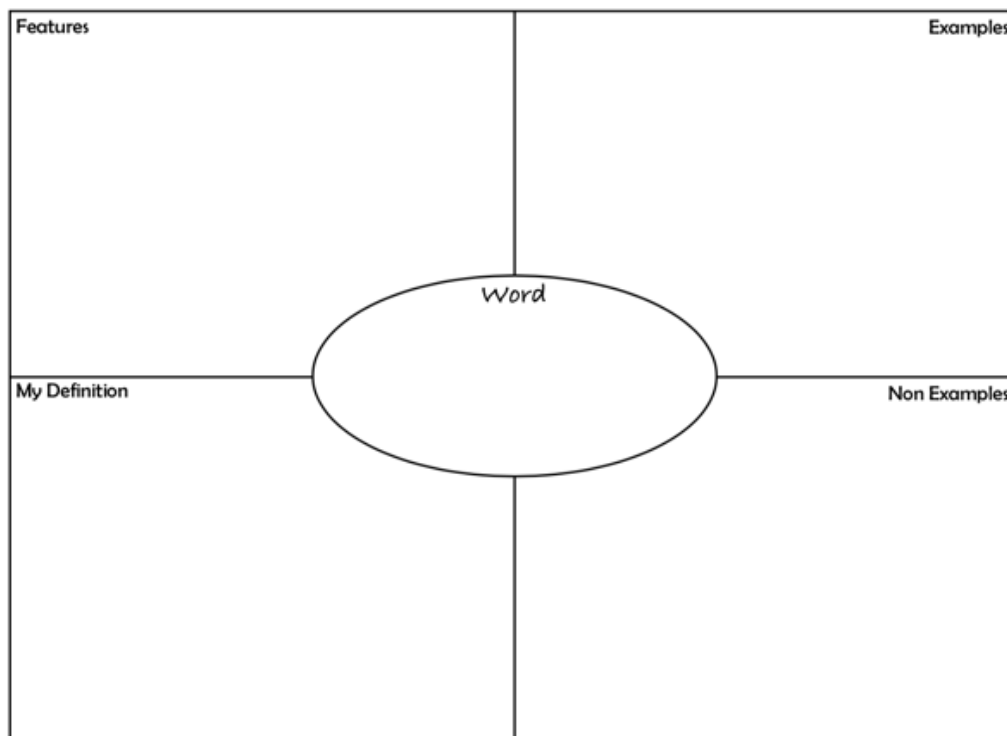
The trial materials were designed to allow participants to receive and reproduce information about the target words in a way that supports concept mastery, as defined by Frayer et al. (1969). The following section describes Frayer et al.'s model of concept mastery as it relates to the study. The process of selecting target words, and creating the instructional videos, the outcome assessments, and other sundry materials used in the study is subsequently described in detail.

#### 4.6.3.1 The Frayer model of concept mastery

Items of vocabulary (words) are labels that we give to concepts. Understanding what a word means, therefore, involves understanding the concept that it represents. Frayer et al. (1969) argue that mastery of a concept (i.e., understanding what a word means) is demonstrated when an individual can: a) name the concept, b) describe its key features, c) give examples of things that conform to the concept, d) give examples of things that do not conform to the concept (i.e., non-examples), and e) define the concept. For example, individuals can be said to understand the concept represented by the word 'quadrilateral' if a) they can name one as such when they see one, b) they can say that a quadrilateral exists in two dimensions, is a geometric shape, has four sides and four angles, is a closed shape, and so on, c) they can identify squares, rectangles, kites, rhombuses, and so on as examples of quadrilaterals, d) they can identify circles, triangles, pentagons, 3D shapes, and so on as

non-examples, and e) they define a quadrilateral as, for example, a type of 2D shape with four sides and four angles.

The Frayer model operationalises this theory of concept mastery into a pedagogical approach. Learners are presented with an item of vocabulary and given key information about the concept it represents. The learners then discuss the item in terms of the five aspects described above (name, definition, key characteristics, examples, and non-examples). Sometimes a teacher guides the discussion directly and sometimes learners are instructed to discuss the item with their peers. The discussion is used to facilitate completion of a 'concept map' (Figure 4.8). Concept maps based on Frayer principles vary in construction, but typically include spaces for learners to write a definition of the item, note the key characteristics of the item, give examples of the item, and provide non-examples of the item.



**Figure 4.8.** Frayer model concept map.

By encouraging consideration of a new item of vocabulary in this way, and by providing a structured frame in which to record the results of that consideration, the concept map makes explicit the types of knowledge that Frayer et al. (1969) argue must be present for a concept to have been mastered. If these different types of knowledge must be present for a concept to have been mastered, then it is reasonable to assume that teaching them explicitly will help achieve this end. In essence, the indicators of concept mastery have been reverse-engineered into a method of teaching vocabulary.

Concept maps are often recommended when teaching vocabulary to EAL learners (*e.g.* Colombo and Furbush 2009, Celic and Seltzer 2012, Echevarría, Vogt and Short 2013) as well as to monolingual learners (*e.g.* Graves and Penn 1986, Gore 2009, Dunston and Tyminski 2013). The approach enjoys empirical support (*e.g.* Peters 1974, Monroe 1997, Nahampun and Sibarani 2014, Palmer, Boon and Spencer 2014) when used with monolingual learners, though some studies have not detected clear effects associated with it (Wixson 1986). While much published advice regarding the use of Frayer model concept maps with EAL learners does not specify which language should be used to partake in the discussion and complete the graphic organiser, the assumption is that both will be done in English. Celic and Seltzer (2012), however, recommend that EAL learners are given opportunities to use their home languages throughout the task. In writing about concept maps, they assert that:

“Incorporating your EBLs’ [Emerging Bilinguals’] home languages in this vocabulary development helps to scaffold your EBLs’ understanding of the new English vocabulary words, and to broaden their academic vocabulary base in the home language.” (Celic and Seltzer 2012:158).

They do not say whether the effects of this approach have been evaluated empirically, and I have been unable to locate any studies that would cast light on this important question. Indeed, it is just this sort of uncertainty about the effects of leveraging children’s home languages that motivated me to conduct this comparison.

#### 4.6.3.2 Introducing the concept maps to the participants

I created exercise books (Appendix I) for each of the participants, which consisted of 22 pages of blank concept maps, preceded by an exemplar that had been partially completed for a common word (dog), not one of the target words used in the study. The first page also included instructions on how to use the concept map.

To ensure that participants were familiar with the process of using a concept map, I introduced them to the procedure before the intervention began. I talked through the instructions on the first page of the exercise book then used the partially completed concept map as a worked example. We discussed possible responses to each segment of the concept map, and worked together to add information to it.

#### 4.6.3.3 Selecting the target words

Target vocabulary items were all superordinate concrete nouns drawn from the English national curriculum for Years 4, 5 and 6. Concrete nouns were chosen because their characteristics are relatively easy to describe. A characteristic of ‘fish’ (a concrete noun), for example, is that it can

swim. A characteristic that is easily identified and communicated. On the other hand, characteristics of 'respiration' (an abstract noun) are harder to unpack and communicate. In addition, concrete nouns are more straightforward to assess in multiple-choice tests of receptive vocabulary (the type of assessment used in this study) because they are easier to represent pictorially. Using the above example as an illustration, if you wanted to assess whether someone knows what the word 'fish' means, you could show them four pictures, only one of which is unambiguously a fish, and ask them point to the fish. Providing a picture that unambiguously depicts 'respiration' is not so straightforward, and the accuracy of the assessment may be confounded by participants' struggle to interpret the picture rather than their knowledge (or lack thereof) of the target word. An assessment of this sort using only concrete nouns is more likely, therefore, to reflect concept understanding, rather than semiotic (mis)understanding.

Superordinates were chosen because the Frayer model requires the learner to give examples (subordinate terms, or what linguists call hyponyms (Kreidler 2014)) of the word being learned. To meet this expectation, the target word must by definition be a superordinate term. To illustrate: if the target word is 'cat' it is relatively straightforward for a learner to find examples (hyponyms) of it - tabby, tortoise-shell, and lion, for example. If the target word is lower down the ordinal hierarchy, say *Siamese Cat*, finding hyponyms is much more difficult ('the antagonistic twin cats in the Disney film *The Lady and The Tramp*', perhaps!?).

To make the word-learning task relevant to the broad purpose of this study (to evaluate alternative ways to help EAL learners do well at school), target words had to reflect the kinds of vocabulary needed to support students' development of their academic literacy. In the context of this study, academic literacy can be thought of as "the reading and writing used in school contexts" (Moore 2008:314), or "the kind of literacy needed for achievement on traditional school tasks and standardized assessments" (Lewis and Reader 2009:105). Academic literacy relies on understanding on a number of planes. Students must understand the features of text genres used in 'traditional school tasks', such as those used in expository writing or lab reports. This understanding in turn relies on understanding the types of vocabulary that are typically used in those genres. These types of vocabulary can be categorised into two broad sub-groups. These are domain-specific academic vocabulary and general academic vocabulary (Baumann and Graves 2010). General academic vocabulary refers to words that are encountered across a variety of disciplines, and they are often quite abstract and difficult to define. Examples include words like *moreover*, *analysis*, *processes*, and *synthesis*. Domain-specific academic vocabulary comprises words that are high frequency within an academic discipline but that are very low frequency in other disciplines or in non-academic texts.

They might be thought of as ‘technical’ words encountered almost exclusively in one subject at school. For example, *mean*, *median*, and *mode* are all examples of academic vocabulary specific to the domain of mathematics. *Humerus*, *tibia*, *fibula* are all examples of academic vocabulary specific to the domain of biology. Because domain-specific academic vocabulary tended to meet my selection criteria of superordinate concrete nouns, where general academic vocabulary tended not to, it was the former that constituted the target words used in my study.

### **Long-listing**

The English national curriculum programmes of study note specific academic or subject-related vocabulary that children are expected to learn. These are not itemised, as such, but are incorporated into the learning objectives for each subject. For example, the programme of study for science in Year 5 states that children must be taught to “describe the differences in the life cycles of a mammal, an amphibian, an insect and a bird” (DfE 2013:168). The superordinate concrete nouns constituting domain specific academic vocabulary in this example are mammal, amphibian, insect and bird. I consulted the programmes of study for English, maths, science, history and geography (DfE 2013). In addition, I consulted past exam papers for the spelling, punctuation and grammar assessment (SPaG test) (DfE 2016b) for other examples of these kinds of words. I found 30 that fit my criteria (see Appendix J).

### **Shortlisting**

The time available for the intervention allowed for the teaching of 18 words; three sessions of three words per phase. In addition, I did not want to include words in the target list that were likely to be widely known to the participants already. To inform the necessary reduction in number, I created assessments of expressive and receptive knowledge<sup>19</sup> of the long-listed words, then field tested these with a class of 31 Year 6, predominantly monolingual English children in a school that did not take part in the study in any other respect. The process was to allow me to identify items of vocabulary that were widely known and to remove them from the long-list of target words.

Items that were correctly identified by at least 80% of the test students in either the expressive or receptive assessments were considered ‘easy’ and removed from the longlist. This generated a shortlist of 20 words. The shortlisted words were included in the baseline assessment used with the actual trial participants, at which point the two highest scoring words were removed to create the

---

19. These were the same tests used to assess outcomes in the main study. See Section 4.6.5.

final set of 18 target words (see Appendix J). Scripts for the voiceovers of the instructional videos were then prepared for each of the target words.

#### 4.6.3.4 Creating and standardising the video voiceover scripts.

I created the instructional videos first by writing English language scripts for each target word. As a primary teacher of 17 years' experience, with extensive experience of teaching EAL learners, I felt confident that I could write suitably pitched scripts for each word. As a check, however, I also consulted online and printed children's encyclopaedias - such as Britannica Kids (Encyclopædia Britannica, Inc 2018) - and books and websites written to support the primary school curriculum - such as the BBC Bitesize series (BBC 2014). I used these as models for the scripts used in the intervention.

The scripts followed the same generic structure based on the information required to complete the Frayer model concept map. Each began with a short definition of the target word, then characteristics of the word were described. Examples of the word were then given, with a prompt to the viewer to try to think of other examples. Finally, non-examples were suggested, again with a prompt to try to think of other non-examples. Below is an example a voiceover script. Similar scripts were created for all target items of vocabulary.

#### ***Script for the word 'Polygon'***

*The word we are learning now is polygon. A polygon is a type of 2D shape.*

*Here are the main characteristics of polygons.*

*Polygons are 2D shapes. They are also called 'plane' shapes. They are made up of only straight lines. The straight lines join up with no gaps. The straight lines are called the 'sides' of the polygon. A polygon must have three sides, or more. A polygon can't have only one or only two sides. Polygons must have three corners, or more. Polygons can't have only one or only two corners. The corners of polygons are also called 'angles'. Polygons can be regular. This means all the sides are the same length. And all the corners are the same size. Polygons can be irregular. This means the sides are different lengths. And the corners are different sizes.*

*Here are some examples of polygons.*

*Triangles are polygons. Triangles have three sides and three corners. Heptagons are polygons. Heptagons have seven sides and seven corners.*

*Can you think of other things that are polygons?*

*Here are some examples of things that are not polygons.*

*A straight line does not have corners. A straight line is not a polygon. A semi-circle has only two sides. One of those sides is not a straight line. A semi-circle is not a polygon. Shapes that are not 2D are not polygons. A cube is 3D. A cube is not a polygon. Things that are not shapes are not polygons.*

*Can you think of other things that are not polygons?*

Once the scripts were written a comprehensibility check was conducted on them to indicate how likely the voiceovers would be understood by an average primary school pupil. As a proxy for comprehensibility of the spoken content of the videos, I calculated the Flesch-Kincaid readability score (Kincaid et al. 1975) for each voiceover script, using the 'readability statistics' feature in MS Word for Mac, version 19.19 (181109). The higher the score the easier the text is to read. Texts that score between 70 and 100 on the Flesch-Kincaid scale are considered 'fairly easy' to 'very easy' to read (Flesch 1979), and, according to the algorithm used by MS Word, texts with a readability score of 70 are considered appropriate for children in their final year of primary schooling (Microsoft 2018). To maximise the comprehensibility of the texts, given the age ranges and EAL status of the participants, a readability score of 80 or greater was sought. The readability calculation was performed, and if a script was not in 80 - 100 range it was adapted, and the readability score re-calculated. This was repeated until the readability scores of all scripts fell between 80 and 100.

Obtaining a readability score within that range was not always straightforward. The Flesch-Kincaid score is calculated by cross-referencing the number of words per sentence with the average number of syllables per word. The academic nature of the target vocabulary means that it is frequently polysyllabic. In addition, the purpose of the text - to define and expound the characteristics of an academic word - meant that these polysyllabic words were often repeated, in many cases twenty or more times over the course of the text. This has a tendency to push the readability score down. Because of these characteristics of the target items, and the effect they had on the overall readability score of each text, they were replaced with monosyllabic place-holder words when calculating those scores. Given the purpose of the text was to explain the meaning of the target words, replacing the target word with a place-holder resulted in a more useful estimation of the readability, and therefore understandability of the text as a whole.

#### 4.6.3.5 Translating and recording the voiceover scripts

Once the English language scripts were complete, they became 'seed' scripts for the L1 translations. Ideally, translations would have been made by certified translators, proficient in academic language in both English and the relevant L1. With an English seed script of approximately 5,500 words and a total of 252 individual videos covering 14 different L1s, use of certified, commercial translators was prohibitively expensive. Bilingual colleagues, friends, and fellow students were approached in the first instance and asked to volunteer their time and linguistic skills to translate the scripts. Using this approach, translations for French, German, Lithuanian, Romanian, Polish, and Swedish languages were obtained.

It proved much more difficult to find local help to translate the scripts into the remaining languages: Arabic, Bengali, Tagalog, Italian, Korean, Panjabi, Spanish, and Urdu. In many cases local users of these languages were unable to dedicate the considerable time and effort needed to do the translations, or felt insufficiently confident to do so. For these, the services of freelance translators advertising on the website [fiverr.com](https://www.fiverr.com) were used. Fiverr is an online marketplace of individual vendors offering low-cost, niche services on a job-by-job basis. The website covers an enormous range of services, many of which are translation services. As an indication of quality, vendors on Fiverr are rated and reviewed by previous customers. Reviews of vendors offering translation into the languages needed for the study were read, and the highest rated translators for the eight languages listed above were engaged to translate the scripts.

The nature of the Fiverr marketplace means that translators are not usually certified. Instead many are simply bilingual individuals looking to supplement their income by capitalising on their linguistic skills. As such they were very similar to the people who volunteered to translate the scripts locally. The nature of this kind of 'informal' translation required oversight to verify the accuracy of the translations. To do this, local proficient bilingual users of the languages represented in the study were recruited. They read both the English and L1 scripts and, after a discussion to clarify any uncertainties about meaning, to make changes or corrections as appropriate. In the vast majority of cases the translations were considered accurate. In the few that had errors, these were very minor and were easily fixed by the bilingual volunteers.

The verified scripts were then audio recorded. The voiceovers for the English language versions of the videos were recorded by the author. The volunteers recruited to verify the translations performed the L1 voiceovers. The audio recording were then prepared for inclusion in the videos.



#### 4.6.3.7 Creating the videos

Each video required images to illustrate the target word. These were either created by the author or found online. Images drew attention to the features of each word, they illustrated examples, and illustrated non-examples. For example, the video for the target word 'reptile' included images of reptile scales and eggs (features), pictures of crocodiles, snakes, and turtles (examples), and pictures of chickens and fish (non-examples). Using the movie making software iMovie, the images were then assembled into a photomontage for each target word. For each video the target word, in English, appeared in writing across the centre of the screen for the duration of the video. Finally, the voiceover soundtrack to the photomontages for each word, in each language, was added. Ultimately 270 unique videos were produced - eighteen for each of the fourteen L1s represented among participants, and eighteen English language versions. Each video lasted approximately two minutes.

#### 4.6.3.8 Making the videos available for the intervention

The completed videos were uploaded to the video sharing site vimeo.com. These were then embedded in a website created using the platform at wordpress.com. On the website, the videos were organised by language and in instructional sets of three (see Section 4.6.4 for an explanation of how these sets were determined). Figures 4.9, 4.10 and 4.11 illustrate the organisation of the website. During the intervention, each participant accessed the videos by visiting the website and selecting the appropriate language and set of words. To protect against possible contamination bias (when members of one intervention group are inadvertently exposed to the intervention intended for members of the other intervention group) (Gorard 2013), the website was password protected so that it could not be accessed outside the intervention session, for example by participants when they were at home. The website can be viewed at [www.vlp2017.wordpress.com](http://www.vlp2017.wordpress.com).

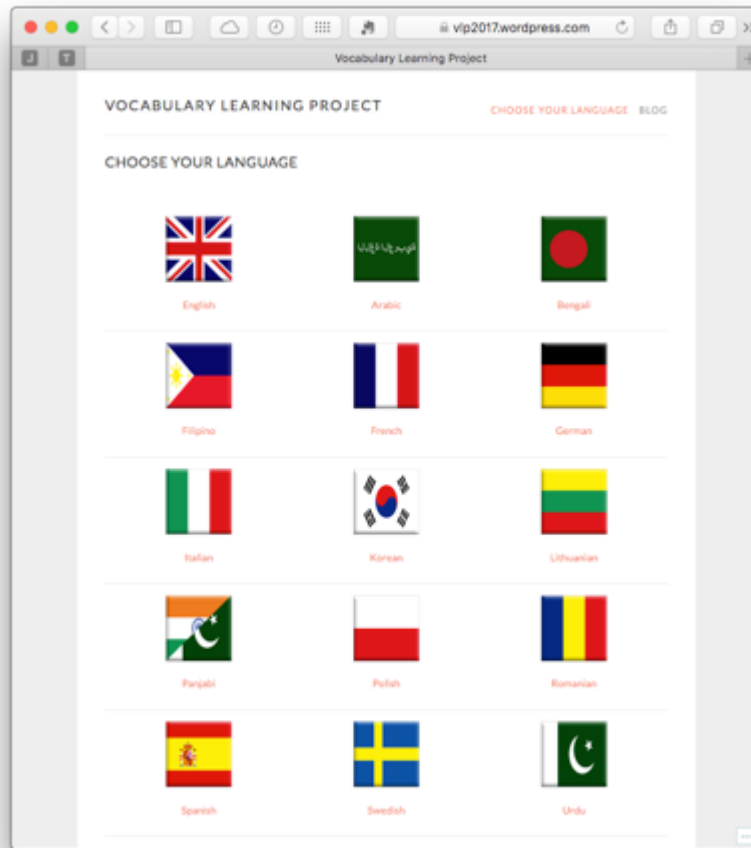


Figure 4.9 Website home page

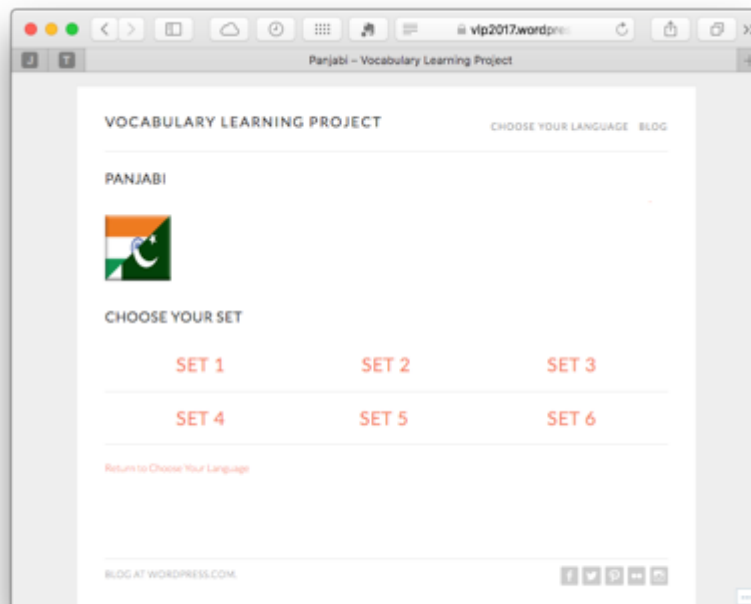
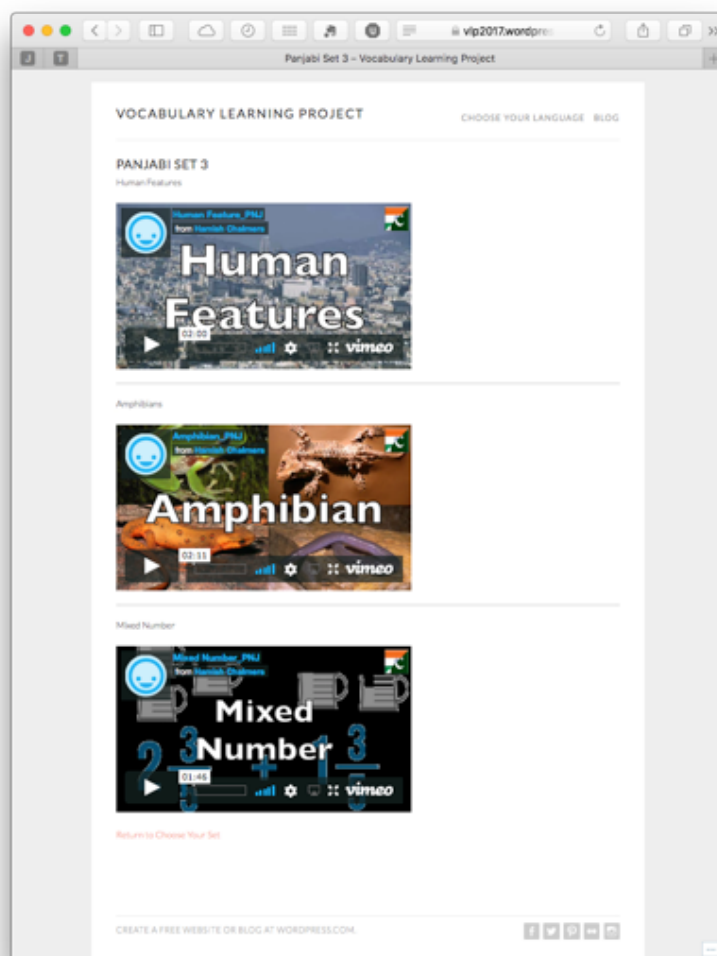


Figure 4.10 Example of a language page on the website



**Figure 4.11** Example of a set of videos on the website.

#### 4.6.4 Instructional order

Each learning session focused on one set of three target items. I had initially intended to organise these sets thematically. For example, I had hoped to have a set of three words from the maths programme of study, a set of three words from the history programme of study, and so on. However, the word selection process did not generate neat groups of three thematically linked words that would allow this. Instead, I used the random sequence generator at random.org to create a randomly ordered list of the words. I then divided this list into groups of three, taking the first three words as Set 1, the second three as Set 2, and so on until I had six sets of three items. Each set constituted the words to be taught in any given intervention session. The order in which these sets of words was studied in each school was determined again using the random sequence generator. Thus, a unique, randomly generated instructional order for the target words was created for each school. A unique, randomly generated instructional order for each participant, while perhaps

preferable, would not have been possible as the post-video discussion relied on all children in the class having studied the same words.

By creating random instructional orders across items and across schools I intended to control for possible non-random variation in the relative difficulty of learning each word. What do I mean by non-random variation in difficulty? Firstly, it is possible that some words are simply more difficult to learn than others. For example, to master the concept of 'a polygon' (one of the target words) one must assimilate into one's existing schema information that may be conceptually more difficult to understand than might be true of the concept of 'a quadrilateral' (another of the target words). To have mastered the concept of a polygon one must understand that polygons: 1) are shapes, 2) exist in two dimensions, 3) are made up of any one of an infinite number of straight lines greater than two, 4) have any one of an infinite number of corners greater than two, 5) are superordinate to a colossal number of hyponyms, most of which the learner is unlikely to have ever encountered (icositetragon, triacontagon, ennecontagon, myriagon, to name but four). By contrast, to have mastered the concept of a quadrilateral one needs to know that quadrilaterals: 1) are shapes, 2) exist in two dimensions, 3) have exactly four sides, 4) have exactly four corners, 5) are superordinate to a restricted number of hyponyms, most of which are likely to be relatively familiar to the learner (square, rectangle, kite, diamond, for example). While the first two items of conceptual understanding in both cases are identical, the remaining three are qualitatively different. For items 3 and 4 the learner must be familiar with the concept of infinity to fully understand these characteristics of a polygon, but need only understand what 'four' means to do the same for a quadrilateral. For item 5, the learner must draw on relatively specialist knowledge to assimilate this information about a polygon, but can fall back on more familiar and colloquial knowledge to assimilate this information about a quadrilateral. It is possible that these differences affect the relative ease with which these two words are learned.

But it is not just the concepts that might contribute to how well they are learned. Other factors may play into the ease of this process. It is possible, for example, that learning one word might be affected by prior study of another word. Consider the target word 'settler'. Might the ease with which it is learned be affected if the participant had earlier studied one of its cognates, the target word 'settlement'. It is possible that characteristics of individual students' home lives might affect how easy they find a word to learn. A child who loves animals and invests his or her energy in understanding and looking after pets might have an easier time learning target words like 'reptile', 'amphibian' and 'mammal' than a child who prefers to play with construction kits. The child who plays with construction kits, in turn, may be better placed to learn words connected to shape and

space, such as ‘polygon’, ‘quadrilateral’ and ‘polyhedron’. It is possible that there are important differences in ease of word learning related to the L1 of the child as well. A child whose L1 is Romanian might have an advantage in learning the word polygon given that the Romanian translation is ‘*poligon*’. Whereas the absence of a cognate relationship with বহুভুজ (*bahubhuj*, the Bangla translation of polygon) may be a relative disadvantage for a child with a Bangladeshi background. It is also possible that the time at which each word was studied relative to the outcome assessment might affect performance on that test. That is, children might score higher on words more recently studied than on ones studied longer ago.

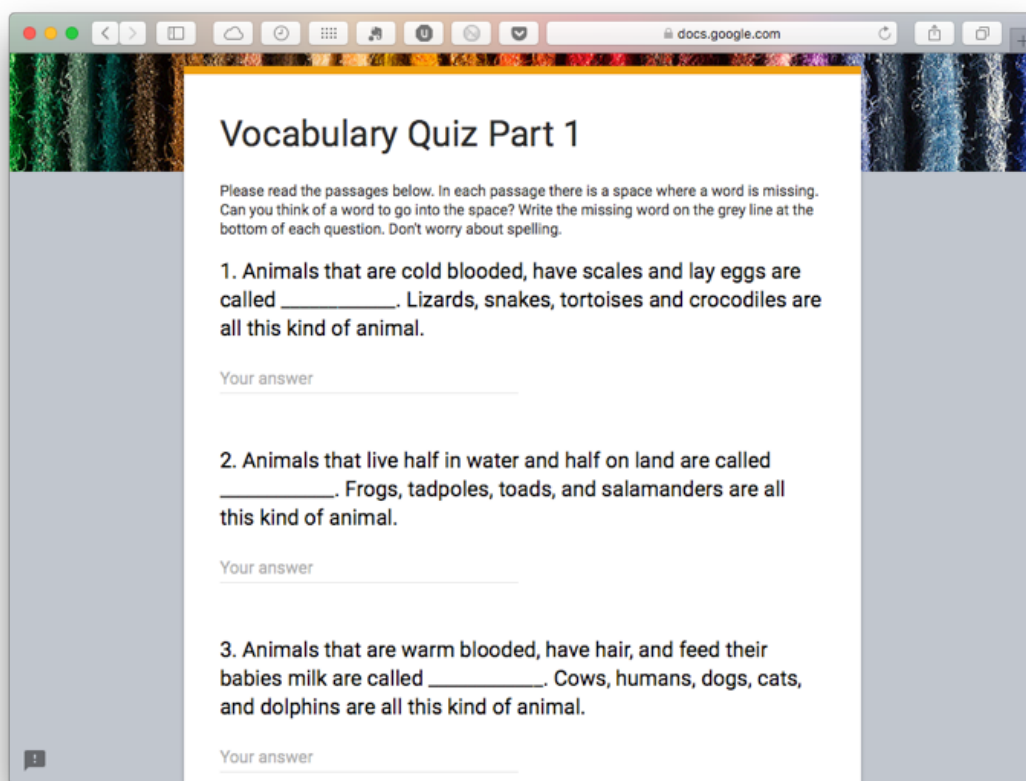
Throughout this section I have made liberal use of the term *possible*. This is not hedging. It is because the accounts I have just given, to greater and lesser degrees, are conjecture. It seems possible that ‘polygon’ might be more difficult to learn than ‘quadrilateral’. It seems possible that a child who likes animals might be more motivated to learn words about animals than a child who likes construction kits. It could be that English words with cognates in a participant’s L1 are more easily assimilated than those with no such relationship. It might be that words studied more recently are better remembered at the time of the test than those learned longer ago. I think I have made a good case for why these are important things to consider. But, I could be wrong. As well as the possibility that I am wrong about the effect of the confounders that I have identified, there might be a million others of which I am blissfully unaware. Without a mechanism to objectively assess the relative difficulty of each target word that takes into account all of the known and unknown potential contributors to that level of difficulty, a randomly generated instructional order is the only sensible option.

The beauty of randomisation in cases like this is that it allows us to be confident that we have done the best we can to distribute target words across the intervention in a way that cannot invite accusations of systematic bias (assigning all of the words with L1 cognates to the L1 condition, while leaving all of the words with no L1 cognates in the English-only condition, for example). While differences in instructional order resulting from known and unknown characteristics of the words may still emerge, they are a result of the play of chance. Given our inevitable ignorance of the extent to which word characteristics might influence the intervention, randomisation is the answer.

#### 4.6.5 Assessments

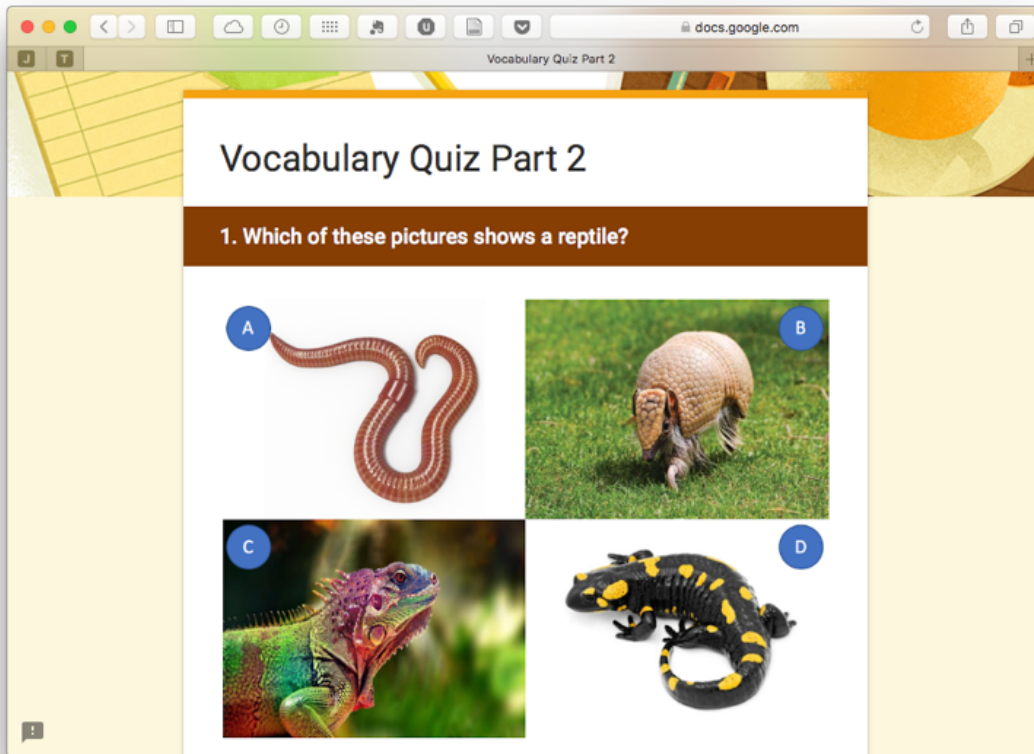
The expressive test was a gap-fill exercise. This was a series of 30 short passages, one for each target word, with the target word missing (Figure 4.12). Respondents were required to fill in the missing

word.



**Figure 4.12** Example of the expressive vocabulary test.

The receptive test was a series of multiple-choice questions. For each item, respondents were presented with four pictures, one representing the target item, and three distractors. The target word was displayed above the picture. Participants were instructed to select the picture that best represented the word (Figure 4.13).



**Figure 4.13.** Example of the receptive vocabulary test.

In both tests, participants were told that if they needed help reading any of the text on the page the researcher could read it to them, but that he could not help them to interpret its meaning.

Tests were created and administered online using Google Forms.

## 4.7 Testing the materials and procedure

The materials and procedures for implementing the intervention and to inform any necessary modification of them were tested with a group of four students who had not been recruited to participate in the main study. This process aimed to assess whether the time allocated to each intervention session was adequate, whether there were any technical issues related to accessing the materials online, whether the Frayer concept map was user-friendly and helped to facilitate discussion about the target words, and whether any other issues associated with implementation arose. The assessment procedures had already been piloted during the word selection phase. No

issues became apparent during the pilot study, and the procedures and materials were retained as designed.

## 4.8 Methods to analyse data

To address RQ1, t-tests were conducted to compare the mean scores in each condition for each type of vocabulary knowledge (expressive and receptive).

To address RQ2, bivariate regression analyses were conducted to assess whether any associations between putative predictors (English proficiency, self-reported L1 proficiency, years in the UK school system, length of residence in the UK, and attitudes to L1) and main outcomes (expressive and receptive knowledge of the target words).

## 4.9 Ethics

The study was conducted in accordance with the ethics requirements of Oxford Brookes University.

### 4.9.1 The ethical warrant for the study

The study sought to address an area of uncertainty about the interventions under investigation, demonstrated by the systematic review presented in Chapter 3. In situations of uncertainty about competing hypotheses, but where practitioners must still make related decisions about practice, the most ethical response is to test the competing hypotheses (Ashcroft 2000). The alternative is to allow policy decisions to be taken on the basis of dogma. When educational decisions are made on the basis of dogma rather than evidence, the potential negative consequences range from wasted time, resources and opportunities at best, to, at worst, active harm to students' development. Given that we do not have sufficient evidence to guide teachers' policy decisions about the use of L1 in L2 schools, and therefore teachers have no option but to take related policy decisions based on dogma, the ethical warrant for the study was met.

### 4.9.2 Informed consent

After headteachers had agreed to participate. Participant information sheets for parents (Appendix C) were sent home with all eligible children in each school. Participant information sheets described the study, and outlined what participation involved, including a statement about their rights while involved in the study. A consent form was appended to this information sheet (Appendix E).

In the meetings held at each school for interested parents and their children, I described in more detail what involvement in the study involved and explained the measures I would take to ensure participants were treated in accordance with Oxford Brookes University's ethics guidelines. Parents and children were given the opportunity to ask questions about the project. I also gave them my



email address and the email addresses of the University's Ethics Officer and my supervisors, to allow them to follow up in private if they wished. At these meetings, I also distributed a simplified participant information sheet for students (Appendix D). The Economic and Social Research Council (ESRC) ethics guidance states that children should have opportunities "to access support in their decision-making, for example by discussing their choice with a trusted adult" (ESRC 2015:29). Therefore, parents of eligible participants were encouraged to talk about the project with their children, and to take their child's wishes into account when deciding whether or not to consent to their participation. Parents of eligible pupils who did not attend the meetings were sent copies of all of this information. All direct written communication with parents was conducted by the school. At no point did I have information about the families' residential addresses, phone numbers or email addresses.

As participants were all legal minors, participation relied on their parents consenting to their participation. This was obtained for all children who participated in the study.

#### 4.9.3 Risk management and data protection

The study was conducted in a climate of mistrust between the Department of Education (DfE) and minority ethnic communities. At the beginning of the school year 2016/17 (the year in which my study was conducted) the DfE introduced a requirement on schools to report the place of birth and the nationalities of their students. This had been criticised (e.g. by pressure group DefendDigitalMe.com) as contributing to the creation of an hostile environment for minority ethnic children and their families. Indeed, the policy was roundly criticised in a *Motion to Regret* in the House of Lords shortly after it was introduced. It was characterised by Lord Storey, who brought the motion, as having "all the hallmarks of racism" and contributing to feelings "of victimisation and being targeted" among members of different ethnic groups (Storey 2016). Despite assurances from the Minister of State for Schools that these data would "be collected solely for internal Departmental use for [...] analytical, statistical and research purposes" and that there were "no plans to share the data with other government Departments" (Gibb 2016), there is evidence that personal information about minority ethnic children was being passed on to the Home Office by the DfE as a matter of routine (Gayle 2016).

In this climate, parents may have been understandably reluctant to be involved in anything that might draw attention to their family's cultural and linguistic backgrounds, and that collected data about their ethnicity, language, and the amount of time they had spent in the English school system. In addition to normal privacy arrangements for studies of this kind, it was extremely important to reassure parents that data collected about them and their children would not be used for any

purpose other than for this study, and to take every reasonable step to ensure that individual children and their families could not be identified from my data, and that data would not be shared with any other organisation (within the limits of legality). It was, therefore, made clear to parents what those steps would be. Each child was assigned a unique identifying number in place of their names, schools were referred to using pseudonyms, data were stored in a password protected laptop computer, backups of those data were stored in a secure online repository, raw data were made available only to the author and supervisors, and, in reporting those data, individual participants could not be directly identified. I also committed to referring any instances of racism or discriminatory behaviour, should I witness it while conducting the study, to the headteachers or heads of pastoral care at the schools, so that they could be considered in relation to the school's anti-discrimination and safeguarding policies. In the event that any participant, their parents, or school wished to discontinue involvement in the study, in line with ethics guidance at Oxford Brookes University, I made clear that they had that right at any point up until the data had been analysed.

#### 4.9.4 Reporting findings to participants

In the field of health research, the moral obligation on researchers to offer participants a summary of the results of the research that communicates clearly and completely what the results say and what they mean, has long been an expectation on researchers (Schulte 1991, Fernandez, Kodish and Weijer 2003, Evans et al. 2011). The educational research establishment in the UK has been slower to realise their ethical responsibility in these terms, with the British Educational Research Association (BERA) stating in their most recent ethics guidelines merely that "researchers have a responsibility to *consider* [...] whether and how to engage with participants at the conclusion of the research" (BERA 2018:8, emphasis added). Having considered this, I committed to making available on conclusion of the study a plain language summary of the results, and my interpretation of the meaning of those results, to any participant who wished to have one. The ethics application was approved by Oxford Brookes University.

# CHAPTER 5

## RESULTS

### 5.1 Introduction

This chapter is organised into three sections. First, RQ1 is addressed through analysis of the effects of the L1 and English-only interventions on each outcome measure (expressive and receptive knowledge of the target words). This is presented as differences in post-test scores by condition, and as the extent of change between pre- and post-test scores in each condition. Next, to address RQ2, associations between participant characteristics and outcomes in each condition are explored through the calculation of bivariate correlations. Based on the findings of that analysis, data are explored to assess whether any significant correlations among sub groups in the sample and outcomes in each condition are differentially associated with the intervention conditions. Finally, the overall findings of the intervention study are summarised.

### 5.2 Restatement of the research questions and methods of analysis

Two research questions were addressed by the intervention study:

RQ1: What are the differential effects of using primary school EAL pupils' L1s to mediate English vocabulary instruction, compared with using only English, on expressive and receptive knowledge of the target words?

RQ2: Are there any statistically significant associations between the level of success on tests of expressive and receptive knowledge of the target words and participants' English proficiency, L1 proficiency, attitude to L1, length of residence in the UK, or time spent in English schooling?

The relative effectiveness of the two conditions (RQ1) was assessed by comparing mean scores on the expressive and receptive tests obtained in each condition. The crossover design is constructed to allow within participant comparisons; the randomised allocation and randomised instructional order of the target words helped to control for possible bias associated with order of instruction and changes over time. Main effects were compared using a paired samples t-tests.

In addition to the average differential effects of the two intervention conditions, the study sought to assess whether particular group characteristics of participants were associated with different outcomes (RQ2). To address this question, bivariate correlations were calculated between theoretically important predictors and observed outcomes. Putative predictors were: teacher assessed level of English proficiency, participant self-report of L1 proficiency, number of completed years in UK schooling, length of residence in the UK, and self-reported attitude towards their L1. Outcomes were expressive and receptive knowledge of the target words taught in each condition. Associations between these variables and scores on the expressive and receptive tests in both conditions were assessed using simple linear regression.

All analyses were conducted using IBM SPSS Statistics version 22.0.0.0. on a MacBook Pro Mid 2014 running OS High Sierra Version 10.13.6

## 5.3 Data Analysis

### 5.3.1 Research Question 1

What are the differential effects of using primary school EAL pupils' L1s to mediate English vocabulary instruction, compared with using only English, on expressive and receptive knowledge of the target words?

#### 5.3.1.1 Descriptive statistics of the results of the post-test

##### ***English-only Condition***

When tested on their expressive knowledge of the nine words taught in the English-only condition, participants scored on average 4.30 (SD=2.65). When tested on their receptive knowledge of the nine words taught in the English-only condition, participants scored on average 6.63 (SD=1.74).

##### ***L1 Condition***

When tested on their **expressive** knowledge of the nine words taught in the L1 condition, participants scored on average 4.25 (SD=2.61). When tested on their **receptive** knowledge of the nine words taught in the L1 condition, participants scored on average 6.48 (SD=2.14). These descriptive statistics are summarised in Table 5.1.

**Table 5.1** Mean scores in each intervention condition

		<b>M</b>	<b>SD</b>
<b>English-only Condition</b>	Expressive knowledge of target words	4.30	2.65
	Receptive knowledge of target words	6.63	1.74
<b>L1 Condition</b>	Expressive knowledge of target words	4.25	2.61
	Receptive knowledge of target words	6.48	2.14

### 5.3.1.2 Distribution

To assess whether the outcome data were normally distributed, their skewness and kurtosis was calculated using the SPSS ‘descriptives’ command. For small sample sizes (fewer than 50 participants), Kim (2013) suggests that skewness and kurtosis values of between 1 and -1 are required to assume normal distribution. As shown in Table 5.2, distribution of scores in all outcome measures were within this range, and are therefore assumed to be normally distributed.

**Table 5.2** Distribution of Outcome Measures

	<b>Skewness</b>	<b>Kurtosis</b>
<b>EO Condition</b>	.023	-.917
<b>Expressive</b>		
<b>EO Condition</b>	-.842	.740
<b>Receptive</b>		
<b>L1 Condition</b>	.040	-.722
<b>Expressive</b>		
<b>L1 Condition</b>	-.911	.209
<b>Receptive</b>		

### 5.3.1.3 Inferential Statistics

Paired-samples t-tests were conducted to compare expressive knowledge of target vocabulary in English-only and L1 conditions and receptive knowledge of target vocabulary in English-only and L1 conditions. Alpha was set at 0.05.

#### ***Expressive knowledge***

No statistically significant difference in scores on the test of expressive vocabulary between English-only (M=4.30, SD=2.65) and L1 (M=4.25, SD=2.61) conditions;  $t(39)=-.16$ ,  $p = 0.88$  was detected.

These results suggest that the language in which definitional information about the English target words is initially provided made no difference to how well those words were subsequently remembered (see Table 5.3).

**Table 5.3** Results of paired samples t-test for expressive knowledge of the target words by condition

	Paired Differences				t	df	Sig. (2-tailed)
	Mean	SD	95% Confidence interval of the difference				
			Lower	Upper			
<b>Expressive knowledge EO and L1 condition</b>	.050	2.025	-.598	.698	.156	39	.877

#### ***Receptive knowledge***

No statistically significant difference in the scores on the test of receptive vocabulary for English-only (M=6.63, SD=1.74) and L1 (M=6.48, SD=2.14) conditions;  $t(39)=-.61$ ,  $p = 0.54$  was detected. These results suggest that the language in which initial instruction about the English target words was provided made no difference to how well those words were subsequently recognised (see Table 5.4).

**Table 5.4** Results of paired samples t-test for receptive knowledge of the target words by condition

	Paired Differences				t	df	Sig. (2-tailed)
	Mean	SD	95% Confidence interval of the difference				
			Lower	Upper			
<b>Receptive knowledge EO and L1 condition</b>	.150	1.545	-.344	.644	.614	39	.543

### 5.3.2 Change in target vocabulary knowledge

Mean differences between pre- and post-intervention knowledge of the target vocabulary items in each condition were compared using paired samples t-tests. These were conducted to assess whether the extent of any change in this knowledge over time was statistically significant, and to assess whether the extent of such change was differentially associated with the two conditions.

#### 5.3.2.1. Descriptive statistics

Average scores before the intervention started (the pre-test) and at the end of each intervention condition (the post-tests) were similar in both conditions for both types of vocabulary knowledge. The extent of the change between pre- and post-tests was also similar. Participants made average gains of between 2.25 and 2.63 points over the course of the intervention in each condition. See table 5.5.

**Table 5.5** Scores on pre- and post-tests of expressive and receptive knowledge of the target words, by condition

	L1 Condition			English-Only Condition		
	Pre	Post	Difference	Pre	Post	Difference
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
<b>Expressive</b>	1.75 (1.69)	4.25 (2.61)	2.50 (1.96)	1.70 (1.67)	4.30 (2.65)	2.60 (1.89)
<b>Receptive</b>	4.23 (1.73)	6.48 (2.14)	2.25 (2.24)	4.00 (2.01)	6.63 (1.74)	2.63 (1.71)

### 5.3.2.2 Inferential statistics

Paired-samples t-tests were conducted to compare pre- and post-intervention scores on tests of expressive and receptive knowledge of the target words in each condition. In all conditions there was a significant difference between the scores on pre- and post-tests. See Table 5.6. These results suggest that participants' knowledge of the target words improved over the course of the intervention and that the extent of that improvement was not a result of the play of chance.

**Table 5.6** Differences between scores on pre- and post-test expressive and receptive knowledge of the target words, by condition

	Paired Differences				t	df	Sig. (2-tailed)
	Mean	SD	95% Confidence interval of the difference				
			Lower	Upper			
<b>Expressive:</b>	2.50	1.96	1.87	3.13	8.06	39	<.001
<b>L1</b>							
<b>Receptive:</b>	2.25	2.24	1.53	2.97	6.36	39	<.001
<b>L1</b>							
<b>Expressive:</b>	2.60	1.89	1.99	3.21	8.70	39	<.001
<b>English-only</b>							
<b>Receptive:</b>	2.63	1.71	2.08	3.17	9.74	39	<.001
<b>English-Only</b>							

Finally, paired-samples t-tests were conducted to compare the differences between pre- and post-intervention scores on tests of expressive and receptive of knowledge of the target vocabulary in each condition. This was done to assess whether conditions (L1 or English-only) were differentially associated with the extent of the gain in word knowledge. No statistically significant difference in the average difference in pre and post scores between the L1 and English-only conditions was detected. These results suggest that the condition under which the target words were learned did not make a difference to the extent of the gain in expressive or receptive knowledge of those words. See Table 5.7.



**Table 5.7** Difference in difference between pre- and post-test knowledge of the target words, by condition

	Paired Differences				t	df	Sig. (2-tailed)
	Mean	SD	95% Confidence interval of the difference				
			Lower	Upper			
<b>Expressive</b>	1.00	2.49	-.70	.90	.254	39	.801
<b>Receptive</b>	.375	2.84	-.532	1.282	.837	39	.408

## 5.4 Summary of findings for RQ1

In summary, the language used for initial instruction was not statistically significantly associated with differences in average outcomes either in tests of expressive or receptive knowledge of the target words. Average scores in both conditions for both outcomes were statistically indistinguishable. Additionally, while participants' expressive and receptive knowledge of the target words improved to a statistically significant degree over the course of both conditions, the extent of that improvement was not statistically significantly different between conditions. Participants made similar gains regardless of the language used for initial instruction.

## 5.5 Research Question 2

*Are there any statistically significant associations between the level of success on tests of expressive and receptive knowledge of the target words and participants' English proficiency, L1 proficiency, attitude to L1, length of residence in the UK, or time spent in English schooling?*

Calculation of skewness and kurtosis for the predictor variables indicated that 'English proficiency' ( $S=-.484$ ,  $K=-.831$ ), and 'L1 proficiency' ( $S=-.031$ ,  $K=-.856$ ), were within the acceptable ranges (Kim 2013) to consider these data normally distributed. However, 'attitude towards L1' at baseline ( $S=-.810$ ,  $K=2.064$ ), 'years in UK schooling' ( $S=-1.052$ ,  $K=.621$ ), and 'length of residence' ( $S=-1.185$ ,  $K=.136$ ) all had values that fell outside this range. Therefore, these data did not meet the assumption of normal distribution.

Because many of these data did not meet the assumption of normal distribution, and because those that did were data at the ordinal level, the non-parametric version of the statistic, Spearman's rho, was calculated for each pair of variables to determine whether any statistically significant correlations existed between these putative predictors and the study outcomes.

### 5.5.1 Correlations

Spearman correlation analysis revealed moderate positive, statistically significant, relationships between English proficiency and expressive knowledge of words taught in the English-only condition ( $R_s=.406$   $p < .01$ ), English proficiency and expressive knowledge of words taught in the L1 condition ( $R_s=.329$   $p < .05$ ), and L1 proficiency and receptive knowledge of words taught in the L1 condition ( $R_s=.326$   $p < .05$ ). No other statistically significant relationships between variables were found. See Table 5.8.

**Table 5.8** Spearman correlations between putative predictors and outcomes

	EO Condition Expressive	EO Condition Receptive	L1 Condition Expressive	L1 Condition Receptive
<b>English Proficiency</b>	.406**	.172	.329*	.149
<b>L1 Proficiency</b>	.260	.248	.165	.326*
<b>Years in UK School</b>	-.159	-.259	-.077	-.182
<b>Length of Residence</b>	-.058	-.100	.033	-.038
<b>Attitude to L1</b>	-.163	-.074	-.008	.012

\*\*Correlation is significant at the 0.01 level (2-tailed).

\*Correlation is significant at the 0.05 level (2-tailed).

In response to finding a positive and statistically significant association between L1 proficiency and receptive knowledge of words taught in the L1 condition, but no such correlation between L1 proficiency and the same outcomes in the English-only condition, a *post hoc* subgroup analysis of the higher proficiency L1 users was conducted to assess whether this group might have benefitted from L1 input relative to English-only input.

Scores on the receptive vocabulary tests for participants who had an L1 proficiency rating of above 3.66 (i.e. in the top third of the L1 proficiency range of 1 to 5) in each condition were compared using a paired samples t-test. No statistically significant differences were detected between outcomes in the English-only condition ( $M=7.26$ ,  $SD=1.20$ ) and the L1 condition ( $M=7.42$ ,  $SD=1.07$ );  $t(18)=-.448$ ,  $p = 0.66$ . High proficiency L1 users considered separately from the full sample, did equally well in either condition on this measure. See Table 5.9.

**Table 5.9** Results of paired samples t-test for receptive knowledge of the target words by condition for only participants with an above theoretical mean L1 proficiency rating

	Paired Differences				t	df	Sig. (2-tailed)
	Mean	SD	95% Confidence interval of the difference				
			Lower	Upper			
<b>Receptive knowledge English-only and L1 conditions</b>	.158	1.537	-.583	.899	.448	18	.660

## 5.6 Summary of findings of RQ2

Calculation of bivariate correlations between putative predictors and scores on post-tests of expressive and receptive knowledge of the target words revealed moderate positive statistically significant correlations between English proficiency and expressive knowledge of words taught in the English-only condition ( $R_s=.406$   $p < .01$ ), English proficiency and expressive knowledge of words taught in the L1 condition ( $R_s=.329$   $p < .05$ ), and L1 proficiency and receptive knowledge of words taught in the L1 condition ( $R_s=.326$   $p < .05$ ). Post hoc analyses of the third of these findings did not reveal any statistically significant relationships between language of instruction and ultimate success on these post-tests that might be attributable to the languages used to mediate instruction in each condition.

## 5.7 Summary of all results

These results suggest that expressive and receptive knowledge of the target words was not associated with the language used to mediate instruction about the meanings of those words. Participants did as well in the L1 condition as they did in the English-only condition. In addition, no associations were found that would suggest that differences in individual participant characteristics (levels of proficiency in English and the L1, length of time in the English school system, attitude to L1, and length of residence in the UK) were associated with the language in which the interventions were delivered.

# CHAPTER 6

## DISCUSSION

### 6.1 Introduction

In this chapter the context and aims of the research project are reviewed and a summary of the literature review is given. The general findings of the systematic review are summarised, then findings of the intervention study are presented and discussed. The implications of these findings are discussed in terms of practice, research and theory. Finally, limitations to the study that should be considered when interpreting the findings and addressed in any replication are presented.

### 6.2 Context and Aim: revisited

The original catalyst for this study was my personal experience as a teacher of primary school EAL learners of an apparent mismatch in assumptions about the role of students' L1s in the context of learning English and learning through English. On the one hand, the SLA research community appears committed to the idea that L1 is a useful mediator of L2 learning. This position is reflected in guides to teaching EAL learners (*e.g.* Celic and Seltzer 2012, Gibbons 2009, Chumak-Horbatsch 2012), which often suggest that judicious and systematic use of the L1 in English medium classrooms results in improved academic outcomes for multilingual students. On the other hand, schools with multilingual students often impose formal or informal 'English Only' policies (Kajonkiet International School 2014, Pattison High School, 2017); teachers (even bilingual teachers) often police the use of L1 among their students (Macaro 2001, Edstrom 2006); and parents and learners often voice concern that anything but exclusive use of English in the classroom will be detrimental to their learning. This apparent collective uncertainty about the role and place of L1 in the L2 classroom goes back at least as far as the 17<sup>th</sup> Century. In 1684 Charles Hoole published a Latin phrasebook that was premised on the idea that "children by the help of their mother tongue may the better learn to speak Latin in ordinary discourse." (Maturin and Hoole 1684). At more or less the same time, schools in which Latin was the prestige language of instruction routinely enforced rules that expressly forbade the use of English (see Watson 1908). The aim of this thesis, therefore, was to address vagueness and uncertainty around L1-mediated teaching and learning approaches in L2 schools, and in particular schools in which many different L1s are represented among the student body.

I have used the term ‘collective uncertainty’ about the place of the L1 in the L2 in setting the warrant for this study. As an overview of the discourse around the issue, it is an accurate description. However, others would argue that ‘uncertainty’ is incorrect, and that, at least in the SLA research community, opinions are remarkably well aligned. Using the term codeswitching to refer to the linguistic practices of multilingual learners, Macaro (2018) makes this plain:

“For the past 20 years or so, the presence of the L1 in the classroom has ceased to become a controversial issue in the SLA community. Anyone attempting to arrive at an overall picture of research and theorizing on the topic will be hard put to find any author arguing in favour of L2 exclusivity. Thus, the argument is no longer among researchers and commentators but among the teachers they are researching. In other words, what we have is a situation such as the following: teachers are (quite rightly) torn in two directions about the benefits of codeswitching. These conflicting feelings may derive from national guidelines, practitioner guides, EFL textbooks, policy statements or from their own beliefs built on years of practice.”

(Macaro 2018:204).

It is thus perhaps more accurate to describe what I have called collective uncertainty as a failure of the SLA community to communicate its understanding of the issue to the practitioners who are the intended consumers of SLA research. Alternatively, it might be that the kind of research produced by the SLA community does not directly address the priorities of educators. The first task in this project, therefore, was to describe the evidence that informs the ‘uncontroversial’ position of the SLA community. This was done through the narrative literature review in Chapter 2, a summary of which follows.

### 6.2.1 Use of the L1: theory and observational evidence

The theoretical literature painted a picture of why L1 is believed to facilitate L2 learning. In particular, Cummins’ (1979, 1980, 2001) construct of linguistic interdependence has greatly influenced thinking around this issue, suggesting that the different languages known to a multilingual individual are mutually supportive, and consequently that linguistic and cognitive development through one language has a positive effect on similar development of and through the other language. Observational evidence supports this conclusion. It has demonstrated that a well-developed L1 is associated with a well-developed L2 (*e.g.* Atwill et al. 2007, Chuang, Joshi and Dixon 2011, Gebauer, Zaunbauer, and Möller, 2013), and has provided evidence that children who receive sustained support in both their L1 and L2 through, for example, bilingual education programmes, perform better than their peers in English-only programmes on broadly defined educational

outcomes such as reclassification from ‘Limited English Proficient’ to ‘English Proficient’ in the American school system (*e.g.* Thompson 2015, Collier and Thomas 2017).

### 6.2.2 Use of the L1: neurocognitive explanations for why this might be helpful to multilingual learners

In addition to theoretical and correlational evidence supporting the position that L1 input positively affects L2 learning, a number of studies have examined how language learners use their L1 when engaging in L2 tasks. Here it has been shown that multilingual learners use their L1 in ways considered to facilitate the L2 learning process. Researchers have argued that L1 scaffolds students’ L2 learning in the following ways: it assists in understanding tasks, improves information sharing between speakers of the same L1, helps to make appropriate linguistic choices in the L2, facilitates higher-order mental processing, and builds knowledge (*e.g.* Swain and Lapkin 2000a, Duarte 2016). By delegating these processes to the L1, it is argued that the cognitive load on the learner is lessened, allowing for more cognitive resources to be directed to the linguistic task at hand (Storch and Wigglesworth 2003).

### 6.2.3 Use of the L1: evidence from experiments

The theoretical and observational evidence provides a compelling basis for exploring whether deliberate teaching in and of the L1 leads to improved outcomes in the L2. However, this kind of evidence considered in isolation is insufficient to support any claim that intentional use of the L1 with multilingual learners will have a causal influence on development of L2. To assert a causal relationship between the two with confidence requires experimental research. The largest, and perhaps therefore, most persuasive body of experimental research in this field compares the effects of bilingual educational programmes with L2-only programmes. Using experimental and quasi-experimental designs, research has found that students who attend bilingual schools do at least as well as their peers in L2-only schools on measures of academic and linguistic performance, and that often they do better (*e.g.* Krashen and McField 2005, Steele et al. 2017a, 2017b).

In addition to research on bilingual programmes there is a small amount of research that addresses the use of L1 as an *ad hoc* approach to teaching multilingual learners. For example, some studies have assessed the effects of allowing multilingual learners to use their L1 when engaging in collaborative tasks (Swain and Lapkin 2000a, Scott and de la Fuente 2008). Some have explored the effects of helping learners to compare and contrast features of their L1 and L2 (Yiakoumetti 2006, Paradowski 2007). Some have explored the effects of using the L1 to give instructions and facilitate engagement in tasks that are otherwise conducted in the L2 (Berman 1994, Bozorgian and Pillay 2013). Many of these, however, were conducted with older learners (not primary school aged), and

overall findings were equivocal. In addition, none of the evidence was obtained from groups of learners representing diverse L1 backgrounds.

#### 6.2.4 Summary of the literature review and next steps

Based on the evidence summarised in the literature review, the position of the SLA community is only 'uncontroversial' if you are prepared to argue that evidence from experiments is unnecessary for asserting causal relationships. With the exception of research on bilingual programmes, there appears to be a remarkable lack of evidence from experiments that addresses use of L1 as a pedagogical tool; even less about how and to what effect it might be used with linguistically diverse groups of learners. To determine whether the narrative approach to summarising the literature had resulted in important evidence being missed, I prepared a systematic review to locate, appraise and synthesise research that addresses directly questions about the effects of L1-mediated interventions with primary aged multilingual learners. A full discussion of the findings of the systematic review is in Chapter 3 (Section 3.6). A summary follows.

### 6.3 Systematic review of intervention studies

#### 6.3.1 Summary of findings

The systematic review interrogated nearly six thousand research reports that keyword searching of bibliographic databases had suggested might meet the inclusion criteria for the review: broadly, empirical evaluations of L1-mediated teaching strategies conducted among primary and pre-primary-aged multilingual children in non-bilingual school contexts, and which reported outcomes as measures of linguistic and/or academic attainment. From this pool of potentially eligible reports, only ten met all inclusion criteria and were incorporated into the final synthesis. Among this vanishingly small body of evidence half of the studies assessed the effects of using the L1 as a mediator for learning L2 vocabulary, the remaining five studies each had a different pedagogical focus, relating variously to reading, writing and speaking. Only one study was conducted with linguistically diverse groups of students. Nearly half of all included studies (four out of ten) did not find evidence that the L1-mediated approach was any better than using only the L2, and two of those studies found that using the L1 resulted in poorer outcomes. Finally, a majority of studies (seven out of ten) were assessed as having high risk of bias, casting considerable uncertainty over the findings of these studies individually and the findings of the review as a whole.

Notwithstanding the general paucity of studies evaluating the substantive effects of L1-mediated teaching strategies, and the lack of experimental rigour, some positive and promising findings emerged in the subsection of the literature dealing with vocabulary learning. Four studies (Seih 2008,

Lugo-Neris et al. 2010, Lee and Macaro 2013, and Camó and Ballester 2015) used the L1 of participants to explain the meaning of target English words in the context of a reading activity, and compared the effect of this with using only English. These all found advantages to using the L1 in this way. However, these studies were all conducted in contexts that differ in important ways to the context in which EAL learners and their teachers in the UK find themselves. For example, the systematic review did not locate any studies that were conducted with linguistically diverse groups of learners (an important characteristic of schools in the UK), there is an imperative to explore whether the findings of the studies that were located translate into similar findings in the UK context.

### 6.3.2 Implications from the systematic review

With such a small and equivocal body of empirical evidence to inform primary school teachers' pedagogical decisions concerning use of L1 for academic and linguistic development in the L2, the tension described by Macaro (2018), (see section 6.2) is understandable. When there is virtually no empirical evidence demonstrating the substantive effects on educational outcomes of strategies that incorporate the L1 to contradict “national guidelines, practitioner guides, EFL textbooks, policy statements or [...] their own beliefs built on years of practice” (Macaro 2018:204) is it any wonder that these latter win out?

The first original contribution of this thesis to the field was, therefore, to have searched for, located, synthesised and summarised empirical evidence relating to the use of L1-mediated teaching strategies in primary and pre-primary L2 contexts, and to find the field wanting. This finding set the warrant for the second original contribution—primary intervention research designed to contribute to addressing the gap in the literature revealed by the systematic review.

## 6.4 Intervention study

The second original contribution to the field was a controlled comparison of a stand-alone L1-mediated strategy amenable to *ad hoc* usage in linguistically diverse mainstream primary schools in England, with an English-only version of the strategy.

Based on advice to teachers of linguistically diverse groups of students (*e.g.* Celic and Seltzer 2012, Chumak-Horbatsch 2012), and informed by the subset of studies that found promise in the use of L1 to support vocabulary teaching with linguistically homogenous groups of students (Sieh 2008, Lugo-Neris et al. 2010, Lee and Macaro 2013, Camo and Ballester 2015), I designed a novel approach to using the L1 to mediate vocabulary learning among linguistically diverse groups of EAL learners in English schools. This approach used pre-recorded instruction about the meanings of key vocabulary items taken from the English National Curriculum, via individual videos watched by each participant.



The instruction was delivered under two conditions: using the L1s of the participants to provide the initial pedagogical input about the target vocabulary and using only English. The differential effects of these two approaches was assessed in a randomised crossover trial, by comparing recall and recognition of the target items taught under each condition. In addition, bivariate correlations were calculated for participant characteristics such as English proficiency and time in the English school system and scores on the post-test, to assess whether these characteristics were associated with differences in outcomes.

## 6.5 Findings of the intervention study

### 6.5.1 Findings for RQ1:

*What are the differential effects of using primary school EAL pupils' L1s to mediate English language vocabulary instruction, compared with using only English, on expressive and receptive knowledge of the target words?*

The study did not detect any statistically significant differences in expressive and receptive knowledge of the target vocabulary between conditions. Regardless of whether the initial pedagogical input was delivered in L1 or delivered in English, average attainment in both conditions was statistically indistinguishable ( $p=0.88$  for expressive vocabulary knowledge, and  $p=.54$  for receptive vocabulary knowledge). This suggests that language of instruction was not associated with subsequent knowledge of the target items.

The study did detect statistically significant positive differences between average pre-test and post-test scores in each condition and for both types of vocabulary knowledge ( $p<.001$  in all cases). The differences were of a similar magnitude and direction in both. Regardless of whether initial instruction was in L1 or in English, participants' expressive and receptive knowledge of the target words improved, on average, by between two and three words in each condition. While this might suggest that participation in the intervention was associated with improved knowledge of the target items, and that the extent of this improvement was not influenced by the presence or absence of instruction in the participants' L1s, no inferences can be confidently drawn about what caused that improvement. Because the analysis relies on simple pre/post comparisons, without a comparison between involvement in the intervention and non-involvement (what might be described as a 'classic control' condition), one cannot say with any certainty what led to that the observed improvement. It is possible that the Frayer model approach to learning vocabulary was responsible. It is just as possible that merely looking at pictures of the target items of vocabulary was responsible. It might be that the discussion (in English) that followed the initial input was responsible for the

instructive effect. Involvement in the study may have made participants more than usually sensitive to the target words when or if they encountered them outside the intervention. It might have been a combination of some, all, or none of these. However, because the extent of the improvement in both the L1 condition and the English-only condition did not differ statistically significantly (i.e. both interventions were associated with a comparable increase in word knowledge), we can be confident that the language of initial instruction was unlikely to have made a difference to the extent of the observed improvement.

The bottom-line conclusion is that it tended not to matter which language was used to mediate vocabulary learning for the EAL learners in this intervention. On average both conditions led to similar outcomes, both in total word knowledge and in relative improvement in word knowledge.

### 6.5.2 Findings for RQ2:

*Are there any statistically significant associations between the level of success on tests of expressive and receptive knowledge of the target words and participants' English proficiency, L1 proficiency, attitude to L1, length of residence in the UK, or time spent in English schooling?*

Bivariate correlations between five putative predictors (English proficiency, L1 proficiency, attitude to L1, length of residence in the UK, and time spent in English schooling) and the four outcomes (expressive and receptive knowledge of the target words taught in the L1 and English-only conditions) were calculated. Three correlations were statistically significant. Two of these were moderate positive correlations between participants' level of English proficiency and their expressive knowledge of words taught in the English-only condition ( $R_s=.406$ ,  $p<.01$ ) and in the L1 condition ( $R_s=.329$ ,  $p<.05$ ). This shows that participants with a better command of English tended to recall more following instruction about English words, irrespective of whether that instruction was mediated in their L1 or in English, than did participants with a poorer command of English. The third statistically significant finding was a moderate positive correlation between participants' level of L1 proficiency and their receptive knowledge of the English words taught in the L1 condition ( $R_s=.326$ ,  $p<.05$ ). This might suggest that higher L1 proficiency leads to improved recognition of English words when instruction is mediated in that L1. However, in the context of the questions driving this thesis, the practical value of this finding for teachers is instructive only if the improvement in vocabulary knowledge mediated through the L1 was statistically significantly different for the same learners when instructed in only English. It wasn't. The findings with respect to RQ2 are explored in more detail in the next section.

### 6.5.2.1 Associations between high English proficiency and recall of words taught in both L1 and English-only conditions

It should come as no surprise that the better a student's command of English, the better placed he or she is to benefit from instruction pertaining to English words, regardless of whether that instruction was mediated by the L1. We are already aware of the positive association between English proficiency and success at school. Analyses of English proficiency and SAT and GCSE scores (e.g. Demie 2018, Strand and Hessel 2018) demonstrates that the better EAL learners know English the better their academic attainment. Nonetheless, it is worth exploring why this might be in the context of this study.

Schema theory (Piaget 1923), underpins the justification offered by Frayer et al. (1969) for their treatment of concept mastery and the subsequent development of Frayer-style concept maps as teaching aids. Schema theory suggests that new knowledge is assimilated into existing networks of related knowledge in the mind of the learner (each network is called a schema, the plural of which is schemata). Cummins' (1980) theory of common underlying proficiency suggests that it does not matter which language is used to incorporate new knowledge into existing schemata, as the underlying proficiency is available to any language used by the learner. Nonetheless, learners with high levels of English proficiency may find that this facilitates assimilation more readily. For example, in this study explicit or even tacit awareness of English morphology may have played a part in the successful learning of new words. Morphological awareness is described by Carlisle (1995:194) as "awareness of morphemic structures of words and the ability to reflect on and manipulate that structure". In both monolingual and multilingual users of English, morphological awareness is known to be positively associated with English vocabulary knowledge (Bowers, Kirby and Deacon 2010; Zhang and Koda 2012, 2013, 2014). This awareness helps learners to understand the meanings or syntactic roles of unknown words (Carlisle, 2003). For example, awareness that the morpheme *-less* indicates the quality of not being in possession of something (*homeless*, for example) helps orient the learner to the potential meaning of unfamiliar words that use the same morpheme (*penniless*, *heartless*, *speechless* for example). When a learner's hypothesis about the new word is reinforced by other instructional components, the word is assimilated into the learner's existing schema for words ending in *-less*. In learners with lower levels of English proficiency it is possible that the relevant schema is less well primed to receive new information and thus the word is less easily learned.

In the context of this study English proficiency must be considered a proxy measure of morphological awareness, as the latter was not measured. Indeed, morphological awareness is used here only as a speculative illustration of why better English proficiency might be associated with better outcomes in this study. The point still stands that there are probably good reasons why a

relationship between higher levels of English proficiency and higher outcome scores exists, and that using the L1 to support learning new words in English-proficient learners may not meaningfully affect these outcomes.

#### 6.5.2.2 In the L1 condition better L1 proficiency is associated with better word learning

It is no less surprising that participants with a well-developed L1 were better able to learn from input in their L1 than participants with a relatively less well developed L1. This finding may not need exploration of abstract psycholinguistic phenomena like schema theory. It may simply be that participants with a better command of their L1 understood what they were being told in the videos better than their peers with lower L1 proficiency. Had there been statistically significantly higher scores in the L1 condition than the English-only condition for participants with a better command of their L1, this might have provided evidence that they were advantaged by L1-mediated instruction. This was not the case, however. The *post hoc* subgroup analysis comparing receptive scores in each condition for participants with above mean L1 proficiency did not reveal any statistically significant differences. Participants with above mean L1 proficiency did equally well in both conditions, reflecting the findings of the study as a whole.

The bottom line is that the five putative predictors are not meaningfully associated with the outcomes in this study. Where there were statistically significant correlations, these did not reveal any information that addressed the question of whether the language of initial vocabulary instruction made a difference to how well the vocabulary was learned. Moreover, with so many correlations being calculated, the chances of finding some that are statistically significant just by chance is likely to be non-significantly different from zero.

#### 6.5.3 A note on variance

Average scores on the tests of expressive and receptive vocabulary were very similar. However, this masks a great deal of variance. The correct answers on the tests of expressive knowledge of the target vocabulary ranged from zero to nine (out of a possible nine) in both the L1 and English-only conditions. For the tests of receptive vocabulary the range was slightly narrower, from two to nine, again in both conditions (with one outlier scoring 1 on the test of receptive knowledge in the L1 condition). No clear pattern that would explain this variance has emerged from the data, as demonstrated by the correlational analysis. Perhaps this is because the number of participants occupying the extremes of the range were so few as to make statistical comparisons unreliable. Nonetheless, this is a curious pattern (or lack thereof). Further investigation of demographic or individual differences among participants in any future replication of this study would be a valuable contribution to understanding why there was such variation in responses to the interventions.

## 6.6 Implications for practice, research and theory

### 6.6.1 Practice – use of the L1 is a personal choice

It is remarkable that faith in L1-mediation as a way to improve L2 linguistic and academic outcomes in EAL primary school children in the UK is so robust given that the body of empirical evidence is so slim and so equivocal.

It should be made clear here that this refers to faith in what amounts to stand alone or *ad hoc* strategies that make use of the L1, as opposed to systemic and sustained instruction in L1; strategies of the sort I have highlighted from literature, aimed squarely at teachers in diverse English schools (e.g. Bourne 2002, Celic and Seltzer 2013). Other arguments for incorporating L1 into the educational experiences of multilingual learners, such as for socio-political reasons, are well made. Multilingual children should not be denied opportunities to use their L1, and explicit celebration of linguistic diversity is a respectable ideological position. Indeed, the *United Nations Convention on the Rights of the Child* states that, among other things, “The education of the child shall be directed to [...] the development of respect for the child’s parents, his or her own cultural identity, language and values” (United Nations 1990:9). The United Nations has consistently reinforced the view that children should be supported in the maintenance and development in their L1s, on socio-political and socio-cultural bases. For example, in an article celebrating International Mother Language Day, the United Nations argued that “all moves to promote the dissemination of mother tongues will serve not only to encourage linguistic diversity and multilingual education but also to develop fuller awareness of linguistic and cultural traditions throughout the world and to inspire solidarity based on understanding, tolerance and dialogue.” (United Nations 2017: no page). More locally, observational research in the UK reports that activities that incorporate the L1 of multilingual children in British schools helps to instil pride in their identities as multilingual members of society (Parke et al. 2002, Kenner et al. 2008).

In the light of the above, rejection of L1 is either politically motivated (for example the demotion of Arabic from its erstwhile position as an official language of Israel (Mandel 2018), or the prohibition on using Kurdish in Turkish schools (Associated Press 2009)), or stated in the belief that using more than one language will interfere with a child’s cognitive development (see Lightbown 2008). The first type of rejection is entirely ideological, so reasoning against it on objective empirical grounds is unlikely to be productive. The second type of rejection is refuted by years of empirical research. We have known at least since Lambert and Tucker’s (1972) study of bilingual education in Canada nearly fifty years ago that multiple languages can be learned simultaneously with no apparent ill effect. Indeed, were it the case that multilingualism is significantly associated with limitations on cognitive

and linguistic development, we might expect to see this reflected in comparisons of educational success between countries in which multilingualism is commonplace and in those in which monolingualism is the norm. No such pattern can be discerned in these kinds of comparison. In fact, in the most recent PISA ranking of countries by the reading proficiency of their 15-year-olds, the countries occupying the top five places are all *de jure* bi- or multilingual societies (Singapore, Hong Kong, Canada, Finland and Ireland) (OECD 2018). Of course, there is more to PISA rankings than the linguistic profiles of the ranked countries that might explain relative differences in educational success in these countries, compared to nominally monolingual countries. Nonetheless, given that more than half the world's population speak two or more languages in everyday life (Grosjean 2010), if concurrent use of more than one language was detrimental to educational success one might expect to see evidence of this appearing more frequently than experience suggests is the case.

The argument for using L1 because it is a beacon of aspiration for a more inclusive and tolerant society, or that it is a medium through which children can fully express their personal identities, are also somewhat ideological. However, no matter the strength of the socio-cultural argument for the use of L1-mediated teaching strategies in UK schools, it remains a proxy argument for the effect that L1-mediated teaching strategies have on academic and linguistic outcomes. In the absence of research that addresses directly the effects of L1-mediated teaching strategies we are in a distinctly impoverished position to argue that case.

The question addressed by this study pertains only to the argument that using the L1 of EAL learners will improve academic and/or linguistic attainment in English. The implications for practice of the findings of this project suggest that, until such time as the evidence changes, teachers of EAL learners in the UK should feel justified in using whichever approach to the L1 they are most comfortable with. The systematic review has demonstrated that considerable uncertainty remains over the effects of using the L1 as a stand-alone strategy (as opposed to more strategic approaches to L1 development such as bilingual education). Moreover, in those cases where L1 use was associated with positive outcomes, the context was very different to those typical of English schools. Most notably, in all but one study, teachers and students shared the same L1, making L1 operationalisation more straightforward than is the case in linguistically diverse schools in England. The intervention study in this project did little to upset the uncertainty revealed by the systematic review. By way of illustration, Figure 6.1 sets the results of the intervention study in context with the other studies included in the systematic review. As can be seen, the present study adds one more to the 'no detected difference' column, moving the body of evidence away from the positive end of the

spectrum, both in terms of all relevant studies and in terms of the subset of studies focusing on vocabulary.



**Figure 6.1** The present study in context with the findings of the systematic review.

The cost to teachers of preparing materials that use the L1, in terms of time, effort and resources, must be considered relative to the potential advantage to students' learning. The time, effort and, in some cases, financial resources needed to write, translate, recruit the help of L1 speakers, record, and otherwise prepare the videos used in the intervention study in this project was extraordinary for such a small set of materials. That it resulted in no perceptible gain for participants suggests that teachers, whose workload pressure is such that the government recently deployed a Teacher Workload Advisory Group (DfE 2018b) to find ways to reduce it, are unlikely to see the effort as worthwhile. This might be offset if commercial companies, or indeed the DfE itself, took on the task of making available multilingual resources for all curriculum subjects in all languages represented among English primary schools. Given the state of the evidence, would this be a financial risk either commercial companies or the DfE would be willing to take? Conversely, if teachers are well equipped to provide L1 mediation, for example if they teach in schools where there is a sizeable majority of

students who speak the same L1, and this L1 is shared by teachers, they should not feel that there is anything untoward about making use of their linguistic resources to provide a multilingual approach to their teaching. They should be aware that this may not result in improved academic outcomes, but that it may meet objectives pertaining to social justice.

From a policy perspective, bodies like the DfE should be more circumspect about advising teachers to make use of the L1s of their students until such time as they can base their recommendations on sound evidence. This point may be moot given the reduction in official guidance from the DfE about EAL in English schools to almost zero following the election of the coalition government in 2010 (Flynn and Curdt-Christiansen 2018). Nonetheless if we are to hope for a return to pre-2010 levels of supporting materials for teachers of EAL learners, we must also hope that documents like Bourne's (2002) *Home Languages in the Literacy Hour* take a more evidence-based approach to the advice they mete out. Teachers need to be confident that the advice they receive from their controlling body is trustworthy.

Finally, on the implications for practice and policy is the possible advantage of establishing paired bilingual programmes where they are viable and desired by parents and students. Bilingual programmes are of course qualitatively very different from use of L1 as an *ad hoc* mediating strategy. In the former, the L1 is not used as a pedagogical tool, as such. Instead L1 and L2 are deliberately taught and developed, and this process is associated with better results on linguistic and academic measures than for children in non-bilingual schools (see Chapter 2). For bilingual programmes to be viable, a "critical mass" (Slavin and Cheung 2005) of students using the same L1 is needed. It is reasonable to assume that in some parts of the UK a critical mass of Urdu/English or Sylheti/English speakers, for example, might exist to allow for bilingual programmes in these languages. However, one must consider the possibility that Urdu/English or Sylheti/English bilingual programmes would result in *de facto* segregation along cultural/linguistic lines. Concerns about *de facto* segregation when children were withdrawn from mainstream schooling for English language instruction in the 1960s and 1970s was one of the motivating factors for the UK government's commissioning of the Swann Report (Swann 1985), which rejected segregation, even for rationalised educational reasons, as "socially and educationally unacceptable" (Costley 2013:283). This finding was reiterated in the later Commission for Racial Equality report (CRE 1986). Together these reports signalled a shift from separation to integration of EAL learners in the mainstream. Arguments have changed, and, at least until the events surrounding the referendum on Britain's membership of the EU in 2016 (on which more later), our understanding of multilingualism in UK society might be said to have matured. It is possible that *de facto* segregation based on linguistic background would be more palatable now than it was in the mid-nineteen-eighties if it can be shown to contribute to closing the attainment gap



evident between EAL learners and their monolingual peers (see Strand, Malmberg and Hall 2015), but this remains to be explored.

## 6.6.2 Research – more and better research is needed

### 6.6.2.1. Volume of relevant research

Clearly more is needed, especially in the UK, if the three-centuries-old collective uncertainty outlined in Chapter 2 is to be settled. The findings of the systematic review in this study is in line with findings from other systematic reviews assessing the extent of the ‘what works’ evidence relating to teaching EAL learners in the UK (Murphy and Unthiah 2015, Oxley and de Cat 2018). Namely, that there is very little of it. Despite this, there is a clear mandate for research that seeks to find ways to close the attainment gap between disadvantaged sections of the school population in the UK (a section that is often occupied by EAL learners) and their more advantaged peers. The Education Endowment Foundation (EEF) was established with a spend down grant of £125 million from the DfE for exactly this purpose (EEF 2012). The willingness to fund ‘what works’ research, embodied by the mission of the EEF but present elsewhere (The Sutton Trust (2012-2018), Unbound Philanthropy (no date) and The Bell Foundation (2019), for example), that addresses inequality in the education system should be seized on by researchers so that policy and practice can be informed by evidence rather than dogma.

The paucity of studies described by the systematic review suggests that researchers interested in the effects of using L1 with linguistically diverse groups of students could pick any area they wished to investigate, such is the extent of the gap in the evidence. Perhaps one of the dozens of untested strategies recommend in *Translanguaging: a CUNY-NYSIEB guide for educators* (Celic and Seltzer 2012) would be an appropriate place to start. This guide, and others like it, have captured the imagination of practitioners the world over. It would be an important contribution to the field if the claims about the strategies recommended in these publications were supported by empirical evidence of their substantive effects on academic and linguistic attainment. Researchers may also like to pick up on the tentative promise demonstrated in the research on using the L1 to teach L2 vocabulary, as was done by the intervention study in this project. The present study notwithstanding, it appears as if the use of L1 to teach L2 vocabulary may be helpful for some groups that fall into the category of EAL. Research of a much larger scale and longer duration, with a more complete representation of the kinds of learner classified as EAL in the UK would build on current evidence, would be well justified, and would contribute important information to our understanding of this specific application of L1 competence.

### 6.6.2.2 Quality of relevant research

It is also clear that, in addition to more research, *better* intervention research on the use of L1 is needed.

Most of the studies meeting the systematic review's inclusion criteria (seven out of ten) were assessed as having a high risk of bias. This calls into question the trustworthiness of the conclusions of these seven studies in particular, and of the review as a whole. When engaging in 'what works' research, researchers should adhere to first principles in their efforts to reduce bias. One of these principles is that comparison groups should be as alike as possible. Randomised trials, in which comparison groups are constructed in such a way as to reduce the influence of systematic differences among participants, should be the design of choice in any new research of this kind. Resistance to the use of randomised trials in education has been vociferous (Deaton and Cartwright 2017, for example), but misplaced. The design is a powerful tool for drawing causal inferences, and is rapidly gaining favour in the UK. Happily, in recent years many more randomised trials have been conducted here than has been the case historically (Connolly 2015). Our understanding of the differential effects of alternative approaches to teaching is the richer for it. Randomised trials to complement the large body of observational research on this topic would be a welcome addition to the canon.

### 6.6.2.3 Clear reporting of research

It matters little how robust the design of a study is or the lengths researchers went to reduce bias in other ways if these are not effectively communicated to the consumers of research. One of the problems in making risk of bias assessments of the studies included in the systematic review here was not necessarily the result of poor practice, but instead poor reporting. A number of studies in the review did not include the information needed to make these judgements. Therefore, adhering to common practice when faced with incomplete reporting (Gorard 2014, B Thomas et al. 2004), studies were given a high risk of bias rating when they failed to provide the information necessary to make an informed judgement. Assigning higher risk of bias to studies where reporting is incomplete or unclear has been criticised (Moher et al. 1995) because it can conflate poor reporting with poor methodology. Nonetheless, the ability to assess the potential for bias in a study is crucial to interpreting the trustworthiness of the conclusions drawn from its results. It is also crucial for assessing whether multiple studies on the same topic can be synthesised in a systemic review.

While it is tempting to find fault with authors for not reporting their studies in sufficient detail to allow readers to make informed judgements, the responsibility for adequate reporting must also lie with the expectations set by journals in the field. Journal editors are responsible for setting the criteria by which reports are accepted for publication. They can, therefore, help to ensure that the

articles they publish conform to the highest reporting standards by insisting that authors adhere to reporting guidelines such as those made available via the Equator Network ([equator-network.org](http://equator-network.org)). The Equator Network website is a repository of more than 400 reporting guidelines, covering virtually every conceivable study design or reporting purpose. These guidelines include CONSORT (Schulz et al. 2010) for randomised trials, STROBE (von Elm et al. 2007) for observational studies, CARE (Gagnier et al. 2013) for case studies, SRQR (O'Brien et al. 2014) for qualitative studies, and PRISMA (Moher et al. 2009) for systematic reviews. Each of these guidelines provides a checklist of essential items that must be included in a research report to ensure that readers have sufficient information about the study to make an informed appraisal of it. While these guidelines were written initially with health research in mind, they are based on first principles common across disciplines. Consequently, they are appropriate for use with education research, perhaps with minor modifications for terminology and to take into account the idiosyncrasies of the field.

Currently, use of reporting guidelines is not common in language education journals. I consulted the 'instructions to authors' pages on the websites of the ten highest ranked British journals in language and linguistics, according to Scimago Institutions Rankings (Scimago Lab 2007-2018). Only one journal (*Studies in Second Language Acquisition*) contained any mention of reporting guidelines. In fact, *Studies in Second Language Acquisition* stands apart as the only journal among the top ten that expressly acknowledges the importance of complete and transparent reporting, stating in its instructions to authors that:

“There is a growing awareness in SLA and in applied linguistics more generally regarding the importance of completeness and transparency in quantitative data reporting practices. In order to maintain the highest possible level of transparency in SSLA, authors are required to adhere to the following guidelines...” (Studies in Second Language Acquisition 2017).

They then list items that must be included in all reports of quantitative research. These include: clearly stated research questions, indicators of the reliability of instruments used, and inclusion of means, standard deviations, and confidence intervals.

By insisting that authors adhere to appropriate reporting guidelines, journal editors can make an extremely valuable contribution to how SLA research is reported and consequently to the confidence with which it can be interpreted and the ease with which multiple studies can be synthesised in systematic reviews. In a field where it is not uncommon to find inordinate preoccupation with the minutia of, for example, referencing style (the journal *Bilingualism: Language and Cognition*, for example, dedicates 811 words to telling its author how to reference other works (Bilingualism: Language and Cognition 2018)), it does not seem out of the question to expect them to include a link

to an off-the-shelf reporting guideline such as CONSORT. If only more journal editors would follow the lead set by their peers at *Studies in Second Language Acquisition* our understanding of the field would be greatly enhanced.

#### 6.6.2.4 Open science practices

In addition to clear reporting, it is widely understood that other so-called ‘open science’ practices improve the quality of research (Manafó et al. 2017, Centre for Open Science, no date). Open science practices include making materials freely available, making data freely available, and pre-registration of studies. Much like the use of reporting guidelines, open science practices are not widely adopted in the SLA community. However, there are nascent moves to improve SLA research by encouraging and rewarding authors for adopting open science attitudes towards their work. The IRIS Database (Marsden, Mackey and Plonsky 2016) collects instruments, materials, stimuli, and data coding and analysis tools used for research into second languages and makes them freely available to other researchers to allow them to replicate existing studies or reproduce them in new contexts. The IRIS database also provides a platform on which to make data from published SLA studies available so that other researchers can analyse them. In the context of this thesis, access to the data from the individual studies in the systematic review may have allowed for statistical synthesis of the results of sufficiently similar studies, rather than the narrative approach that was adopted in the absence of full data reporting. In addition, the availability of data from these studies may have allowed other analyses to help generate a more nuanced picture of the effects of L1-mediated vocabulary teaching among different groups. The third open science practice, preregistration of reports, is also widely understood to improve research by allowing journal editors, peer reviewers, and readers to detect data dredging and outcome switching. These poor practices can happen when researchers do not find an interesting result when they test for their prespecified outcomes, so instead perform lots of statistical tests until they find one that is interesting. They then report the ‘dredged’ outcome and not that which the study was designed to investigate. By doing this they increase the likelihood of a false positive (Goldacre et al. 2019), and our understanding suffers as a result. The SLA journal *Language Learning* is spearheading the movement to encourage open science practices in SLA research by rewarding authors who adhere to these practices (Language Learning 2019). One must hope that other journals recognise the value of open science and follow suit.

#### 6.6.3 Theory - A problem with the orthodoxy?

The results of the intervention study and the direction of findings in the systematic review introduce a pebble into the sandal of the SLA research community. If, as tends to be claimed, correlational relationships between L1 and L2 are actually causal relationships we would expect to see evidence of this in experimental research. That no such relationship was detected in the intervention study,

and uncertain evidence was revealed by the systematic review, suggests that this claim is overstated.

Correlational and other observational data considered in isolation are insufficient to support any claim that intentional development of L1 will have a causal influence on development of L2. It may be that the associations observed between L1 and L2 proficiency described in Chapter 2 are the result of a third 'lurking' variable, perhaps an innate learner characteristic. A useful thought experiment on this possibility can be conducted using Carroll's (1962) construct of language aptitude. Carroll defines language aptitude as "a specialized talent (or group of talents), relatively independent of those traits ordinarily included under "intelligence"" (Carroll 1962:89). Sparks and Ganschow (1991:11) build on Carroll's definition to suggest that second language aptitude "may be strongly related to one's *native* language aptitude" (original emphasis). If we were to test the hypothesis that a learner trait like aptitude is responsible for the positive correlations observed between L1 and L2 proficiency, what evidence might we accept as indicative of its truthfulness? Presumably, exactly the same observational evidence that has been used to argue that development of the L1 has a positive causal influence on development of the L2. When precisely the same data can be used to argue for the veracity of both positions, singling out language as a special case for transfer seems redundant. Positive L1/L2 correspondences might as easily be the result of learner traits, as characteristics of their educational experiences.

Cummins (2016) reports hearing similar criticism of his linguistic interdependence hypothesis. He gives an example: "Can the relationship between L1 and L2 reading comprehension be explained by the fact that both are related to underlying cognitive attributes of the individual?" (Cummins 2016:5). He responds by asserting that attributes and transfer are two sides of the same coin. "Attributes (e.g., verbal cognitive abilities)" he argues, "develop through experience; in other words, they are learned. Once they exist within the individual's cognitive apparatus [...] they are potentially available for two-way transfer across languages" (Cummins 2016:5). If the sociolinguistic and educational conditions are right, he says, transfer of those attributes will occur.

This explanation does not address the criticism implied in the question. Cummins is just restating the basis for his theory of cognitive underlying proficiency: that cognitive attributes reside in an engine that drives performance in each of the languages known to the user. He asserts that these attributes are learned, which is undoubtedly true for some, but ignores the possibility that others might be innate, as I have suggested might be the case for language aptitude. Accordingly, Cummins fails to provide a reason not to conclude that positive relationships between L1 and L2 are actually causal relationships between each of them and a third variable. The pedagogical implications thus subtly

shift from the importance of maintaining the L1 to the importance of providing opportunities to tap into the underlying trait.

The findings of this study for theory suggests that it needs modification. The theory of linguistic interdependence is underspecified in as much as it asserts a relationship between L1 and L2 but offers little by way of explanation of the mechanisms at play in that relationship. Cummins' vague assertion that skills and transfer are two sides of the same coin does not clarify that underspecification. The challenge to theorists is to review theory in light of the findings of experimental research, slim though they are, so that a light might be shone into the black box of common underlying proficiency.

## 6.7 Limitations of the study

There were a number of limitations to the intervention study part of this project. Interpretation of the findings of the study should be undertaken with these limitations in mind. Any replication of the study should seek to address them.

### 6.7.1 Recruiting a representative sample of the population of EAL learners

One of the challenges faced in this study was recruiting participants for whom the intervention was theoretically likely to help. The axiom among the SLA research community tends to be that L1-mediated learning is beneficial to all multilingual learners, regardless of their levels of proficiency in both or either the L2 and the L1 (Cummins' threshold hypothesis notwithstanding). García is clear in her introduction to Celic and Seltzer's (2014:5) *Translanguaging: a CUNY-NYSIEB guide for educators*: "All students would benefit from the translanguaging instructional contexts and strategies offered in this guide." Nonetheless, the empirical studies that directly informed the present study (Sieh 2008, Lugo-Neris et al. 2010, Lee and Macaro 2013, Camó and Ballester 2015) were all conducted with participants who were competent users of their L1s but who were relatively new to English. The studies by Sieh (2008), Lee and Macaro (2013), and Camó and Ballester (2015) were conducted in the context of English as a foreign language classes in mainstream schools in the participants' home countries. Lugo-Neris et al.'s (2010) study was conducted with pre-school learners of English who lived in linguistically homogenous Spanish-speaking communities in the USA. Thus, in a study designed to assess whether a similar approach with EAL learners in English schools would reflect similar advantages, at least part of the sample ought to have been students with a similar profile of linguistic proficiency. Recruiting students with this type of profile was less successful than recruiting students with high levels of English proficiency, though all were classified as EAL learners.

Recall that the term EAL covers everything from complete novices in the English language, through balanced English/L1 bilinguals, to English-dominant learners who are exposed to another language at home but who don't necessarily use it. When parents were invited to consent to their child's participation, it was hoped that a broad spectrum of responses would be forthcoming, from parents of children with low levels of English to those with comparatively well-developed English language proficiency. As is clear from the descriptive data about those who ultimately participated in the study (see Chapter 4), the sample was asymmetrical in favour of children with relatively high levels of English. More than half ( $n=22$ ) were classified as either 'Fluent' or 'Competent' by their teachers. Only seven participants (17.5%) were classified in the lowest two proficiency bands ('New to English' and 'Early Acquisition'). The following section discusses possible reasons for the study's failure to recruit from across the spectrum of EAL learners.

#### 6.7.1.1 'Hard to Reach' Parents

One issue with seeking to recruit participants with low levels of English and contrastively higher levels of L1 is that their parents may be what is sometimes described as 'hard to reach' (Feiler 2010, Lueder 2011). In the case of parents of EAL learners this can be because of low levels of English proficiency, unfamiliarity with the English school system, and differences in cultural expectations on the home/school relationship. These types of characteristic are often represented among the parents of EAL learners, especially those 'New to English' (Ofsted 2004, Husain 2018). Add to this the often quite detailed and sometimes daunting information sent to parents in requests for participation in school-based research, and one can see how the children for whom the intervention might benefit most are easily lost to participation.

Crozier and Davies (2007) take issue with the term 'hard to reach', saying that it "pathologises the parents, laying the blame on them ... [and serving] as an excuse by schools for not being more proactive" (Crozier and Davies 2007:296). In this study, proactive measures were taken to engage with the parents of eligible participants. Going beyond a simple letter of invitation, parents were engaged at the school gate, parent information meetings were held both at the beginning of the school day at the end to coincide with parents' presence on the school grounds, and parents were reminded by email to return consent forms. Even so, this approach was not as successful as might have been hoped. The possible consequence of this is that potentially important characteristics, such as English proficiency, of those who agreed to participate did not reflect the actual diversity across all eligible participants. Unfortunately, failure to recruit (and thus failure to obtain permission to see pupil data for non-recruited students) means that there was no opportunity to assess if indeed particular pupil characteristics were underrepresented. It seems likely though.

### 6.7.1.2 EAL, stigma and immigration

In addition to the challenge of reaching parents with low levels of English literacy, or who, for whatever reason, were less likely to engage with their child's schooling, was the visceral attitude towards immigrants and immigration that characterised much of the political discourse in Britain at the time of the study. The referendum on Britain's membership of the European Union had been held a few months before recruitment to the study began. Concerns over immigration before and since the vote dominated public discourse, with 48% of those polled at the time of the referendum citing immigration as the most important issue facing Britain (Ipsos/MORI 2016). The most egregious examples of rhetoric around this issue employed discourse and iconography resembling the Nazi propaganda of the second World War (see Stewart and Mason 2016). It portrayed immigrants as the architects of Britain's imagined downfall and gave succour to those members of our society for whom this kind of thinking was either germinating or well established (Harris 2018).

In this climate, at the beginning of the school year in which the study was conducted, the Department for Education introduced the requirement for schools to seek information about the nationality and country of birth of their pupils. Overreaching the legal requirement to merely ask parents what the nationality and country of birth their children was, some schools took it upon themselves to demand that parents present birth certificates, passports and immigration documents (Whittaker and Camden 2016). The policy, even without the overreach of some schools, was an extension of an act of parliament (Immigration Act 2014) designed specifically to create what the then Home Secretary described as "a really hostile environment for illegal migrants" (Travis 2013). The hostility may have been aimed at illegal migrants, but it did nothing to reassure legal migrants that they would not be caught up in a general anti-immigrant stance. Indeed, as was demonstrated during that period by the erroneous deportation of members of the so-called Windrush generation (Commonwealth citizens who had legally migrated to the UK between the late 1940s and early 1970s), legal immigrants were clearly not immune from the Home Secretary's willingness to "deport first and hear appeals later" (May 2013). That the Home Secretary chiefly responsible for the hostile environment policy had ascended to the position of Prime Minister by the time of the study will not have reassured families with a history of migration that anything was about to change.

The policy of collecting nationality and country of birth data was met with resistance from pressure groups such as Defend Digital Me ([defenddigitalme.com](http://defenddigitalme.com)) and Against Borders for Children ([schoolsabc.com](http://schoolsabc.com)). They argued that collection of such data was in contravention of the human rights of the child, and encouraged parents to boycott the school census by refusing to provide this information (see Adams and Belam 2016). In this febrile atmosphere it is perhaps unsurprising that the vast majority of parents with eligible children (162 of 208 invited) did not respond to invitations



to participate in an investigation that would inevitably draw attention to the fact of their children's connections to other parts of the world.

### 6.7.2 Selection bias

A general failure to recruit a larger, more representative sample from those eligible to participate almost certainly resulted in selection bias and volunteer bias (Catalogue of Bias Collaboration 2017a, 2017b). Selection bias is when the sample participating in the research do not represent the population being studied because of either conscious or accidental bias in the way they are selected to participate. For example, teachers may encourage only the children they feel are likely to benefit from an untested intervention to participate. In this study, for example, a child whose parents had agreed to his participation was removed from participation by his school because, in their view, he 'wasn't right for the study'. Volunteer bias is a type of selection bias that results from systematic differences between those volunteering to participate and the population of study as a whole. As the introduction to this thesis reported, there are strong and polarised opinions among stakeholders about the value of the L1 in learning English. If parents subscribe to the widely held belief that interventions that use the L1 are ineffective and possibly damaging it is easy to see how they might conclude that volunteering would not be in their child's best interests. Indeed, a parent of one eligible student made contact to say that her children were 'too good at English' to be of any use to the study and so she would not be consenting to their participation. By the same token, parents who subscribe to the view that the L1 has an important place in their child's education, who are better disposed to multilingualism, and who have already invested effort in maintaining and developing the L1s of their children, may have been more likely to volunteer. The implication of these possible biases is that only participants who were unlikely to benefit from the L1 intervention were recruited.

Problems associated with a failure to recruit participants for whom an intervention is theorised to be most likely to benefit were satirically highlighted by Yeh et al. (2018) in a paper in the British Medical Journal (BMJ). The BMJ has a tradition of publishing tongue-in-cheek articles for its Christmas edition. In 2018 it published a report of a randomised trial comparing the effects of wearing a parachute while jumping from an aircraft with wearing an empty backpack. Recruitment took place at two sites, in a passenger airliner and a biplane. Passengers on these aircraft were asked if they would be willing to participate in a trial in which they would be randomly assigned to wear a parachute or an empty backpack while jumping from the aircraft in which they were sitting at the time of recruitment. Ninety-two people were eligible for participation, 23 of them agreed to participate and were randomised. The study found that parachutes did nothing to reduce instances of death and major trauma ( $p > 0.9$ ). These findings might be explained by the baseline characteristics of the participants. Importantly, at the time of recruitment all those who agreed to participate, and

who were thus recruited, were sitting in a grounded stationary biplane. Those who did not agree to participate, and who were therefore not recruited, were all sitting in an airliner traveling at 800 kilometres per hour at an altitude of 9146 metres.

One purpose of the article was to remind us that context is everything in interpreting the results of a randomised trial. Another, but no less important, purpose was to highlight the problem of selection bias. As the authors say, the null finding:

“...largely resulted from our ability to only recruit participants jumping from stationary aircraft on the ground. When beliefs regarding the effectiveness of an intervention exist in the community, randomized trials evaluating their effectiveness could selectively enroll individuals with a lower likelihood of benefit, thereby diminishing the applicability of trial results to routine practice.” (Yeh et al. 2018:5).

In the context of the present study, the pre-existing beliefs among the target population about the risk/benefit of taking part in the investigation may have resulted in the (self)selection bias evident in the sample.

If this second group constitutes Yeh et al.'s (2018:5) “individuals with a lower likelihood of benefit”, our ability to interpret the findings of the trial as it applies to the target population (all children classified as having EAL) is limited.

One way to address the issue would be to review our attitudes to the ethics of controlled comparisons in education, and in particular the potential to bias results by insisting on opt-in consent for individual students when gatekeeper consent has already been granted. That is, where the headteacher has reviewed the justification presented by the researchers for the trial and has agreed to allow the research to be conducted in her school. Connolly et al. note that the “constant recycling of stylised objections” to randomised trials “betrays a fundamental, worrying and widespread ignorance among significant sections of the educational research community of the use of RCTs in reality” (2017:8). One of these ‘constantly recycled objections’ is that it is unethical to expose some children to a promising new teaching approach and to deny it to others, hence the requirement for opt-in consent as one way for parents to ensure that their child is not exposed to inferior teaching. The ignorance betrayed here is the belief that new approaches are inevitably better than existing approaches. If researchers know that a new approach is superior to an existing approach, then they have no business conducting a trial. In the certain knowledge that one approach is better than another, it would indeed be unethical to deny that approach to some participants. If, however, there is uncertainty about the relative effects of a new approach, then a well-designed controlled comparison is the only ethical way forward. The alternative is to continue to teach using approaches

for which we have no or uncertain evidence, quite possibly to the disadvantage of the students. The perverseness of this ethics standard is laid bare when one considers the typical ways new teaching approaches are introduced in schools. Teachers routinely decide what new (often untested) interventions they will adopt in their teaching, without asking for consent from their students' parents. In one sense, therefore, they are always experimenting as they adopt a trial and error approach to their pedagogic choices. In the context of a randomised trial, the only substantive difference is that trialists carefully assess the effects of the intervention against an alternative so that they can learn about their differential effects. If headteacher 'consent' is sufficient to allow teachers to adopt whatever intervention they see fit with no formal evaluation, why is this standard not applied to interventions that *will* be carefully evaluated?

This "double standard on informed consent" (Chalmers and Lindley 2001) was captured by paediatrician Richard Smithells more than four decades ago, observing "I need permission to give a drug to half of my patients but not to give it to them all" (Smithells 1975). We would be in a much stronger position to learn more about our theoretically justified approaches to teaching if we were prepared to allow headteachers to decide whether *the school's* participation in a trial is justified, like they do for virtually every other pedagogical decision taken in their name. Trials would enjoy better statistical power and be less prone to selection bias of the type described above. We would all be the richer for it.

In any future replication of the trial, every effort should be made to recruit participants in the lower bands of English proficiency ('New to English' and 'Early Acquisition') so that more detailed, and hopefully more illuminating, analyses of subgroups within the EAL population can be conducted.

### 6.7.3 Accurate language proficiency assessment

Another limitation to the study was the potential for problems with the validity and reliability of the language assessments in both English and L1. The usefulness of the regression analyses conducted to address RQ2 relies on accurate information about participants' language proficiency.

#### 6.7.3.1 Proficiency in English

As I have described above, many of the participants were assessed as having high levels of English proficiency. In addition, based on my interactions with the students over the course of the study, it appears that some of those in the lowest bands had either been misclassified or had made significant progress since classification. The legal requirement to assess and report English proficiency of EAL learners had been somewhat sprung on teachers by the DfE at the beginning of the academic year in which this study was conducted. Teachers were instructed to conduct these assessments, with no support or training for teachers unfamiliar with assessing language proficiency

and little by way of official guidance. The assessment was rushed; initial assessments were required before the end of the first term of the school year, and reporting these data to the school census was required by February of term two (Scott 2016). Robust judgement of language proficiency is underpinned by thoughtful and sustained formal and informal assessment across a variety of domains (The Bell Foundation 2017). However, contrary to this principle, teachers were given a set of five descriptive criteria that did not differentiate between the four domains of language (reading, writing, speaking and listening) and asked to match their EAL learners to one of them. Even if teachers were suitably qualified to make judgments about the English proficiency of their students, such a short timeframe and such vague criteria against which to judge proficiency, will have likely compromised the validity of those judgments for any who had not been routinely keeping language proficiency data well in advance of the new requirement. For many, judgements will have been based on incomplete data and 'gut feelings'. For example, it is possible that recent immigration was used, erroneously, as a proxy for low levels of English for some children. The two participants assessed as in the lowest proficiency band ('New to English'), for example, were new to their school at the beginning of the academic year in which the study was conducted. Both, however, demonstrated a command of English that did not reflect this judgement. In fact, one of these students revealed that she had attended an English medium school in her home country before moving to the UK, suggesting that, while new to the UK, she was not 'New to English'. It is perhaps understandable that teachers, under pressure to report assessment data about children whom they had only just met and about whom all they knew was that they had recently arrived from overseas, will have erred on the side of conservatism in their judgements. Nonetheless, this vignette underscores the potential issue of validity in these judgments.

As well as the validity of the method of assessing proficiency in English, the reliability of the approach can also be questioned. Matching children to vague attainment descriptors that collapse all skills into a single paragraph has been criticised both from within the DfE and by researchers in the field. The National Curriculum level descriptors, which were used in a very similar way to the English proficiency levels, were described by former Parliamentary Under-Secretary of State at the DfE, Elizabeth Truss, as subjective, lacking in validity, and insufficiently detailed to accurately describe what children are able to do (GOV.UK 2014). Robert Coe of Durham University's Centre for Evaluation and Monitoring characterises them as "meaningless numbers based on unstandardized, impressionistic, selective and biased judgements that fail to capture true learning" (Coe 2014). If, as seems likely, the English proficiency scale is subject to the same frailties as the National Curriculum levels, then confidence in their reliability as a measure across schools can be legitimately questioned.

I stand by the use of these level descriptors in this study to the extent that it has always been a study grounded in the day-to-day realities for teachers. Teachers rely on data from their own and others' judgements about students to inform their pedagogical decisions. To have ecological validity, and therefore to help in making informed choices, the study had to reflect real life for teachers. Nonetheless, issues with the validity and reliability of the assessment tool compromises interpretation of the subgroup analyses that used English proficiency as a predictor. The study is therefore less well placed to speak to theoretical questions associated with linguistic interdependence, or to suggest to teachers whether L1-mediated instruction is more or less useful for students at different levels of English proficiency. In any future replication of this trial an objective, validated measure of participants' English proficiency should be taken at baseline so that a more confident assessment of any related differential effects associated with the measure can be made.

#### 6.7.3.2 Proficiency in L1

Accurate assessment of L1 was even more problematic. Internally valid and reliable tools to assess the L1 proficiency of participants in all 14 L1s, which were also comparable across languages were unavailable, as were individuals with the requisite knowledge of these languages to administer such assessments. The study relied on participants rating their ability to use their L1 on a five-point Likert scale. This gave a reasonable idea of how well the participants felt they were able to use their L1 but no way of confirming the accuracy of that self-assessment. Moreover, the subjective nature of this self-assessment means that it cannot be assumed that ratings from participant to participant were comparable. A rating of 5 on the Likert scale from one participant, for example, does not necessarily reflect the same level of L1 proficiency as a rating of 5 from a different participant. To help address RQ2 in any replication of this study, and indeed any future study that hopes to assess whether L1 proficiency predicts outcomes in L1-mediated learning among linguistically diverse groups, reliable measures of L1 proficiency that are valid across the different languages represented are needed.

#### 6.7.4 Summary

In response to claims that the L1s of EAL learners can be leveraged to support linguistic and academic attainment in English, this project found that very little evidence appears to exist to support that claim. When put to the test in a controlled comparison, L1 did not confer an educational advantage relative to use of only English. These findings suggest that teachers should continue to adopt whichever pedagogical approach to the L1 of their students that they are most comfortable with. In addition, theories of linguistic interdependence should be reviewed to take into account what can be legitimately claimed on the basis of empirical evidence.

# CHAPTER 7

## CONCLUSION

The purpose of this study was to respond to advice directed at teachers in linguistically diverse English primary schools to use strategies that incorporate the L1s of their EAL learners into their teaching 'method'. Such advice is often accompanied by explicit and unequivocal claims that it will improve multilingual students' academic and linguistic outcomes. For example, "Supporting home languages [...] increases children's overall levels of attainment." (Bourne 2002:73). The nature of this kind of advice is problematic.

The first way in which this kind of advice is problematic is that it can be so vague as to have little instructive value for teachers. It is common to find references to 'systematic', 'selective', 'judicious', or 'principled' use of the L1 in second language teaching (*e.g.* Cook 2001b, Butzkamm 2003, Moore 2013), with little to illustrate what these terms look like when applied in the classroom. Littlewood and Yu (2011:75) capture this problem succinctly: "The terms 'systematic', 'selective' and 'judicious' occur again and again in discussions [about L1 use in L2 classrooms] but what do they really mean? At present they have to be interpreted mainly through an individual teacher's intuition." Individual teacher intuition is a valuable resource, and teachers spend their careers accruing and honing the skills, experience, and insights necessary to inform their intuitions. One might expect, therefore, a more experienced teacher to have a more keenly developed sense of intuition than that of a relative novice. But if we want language teaching to be an evidence informed undertaking - where the quality of a learner's education is not contingent solely on their teacher's experience and insights - relying on intuition alone falls short of the ideal. Moreover, relying on intuition alone exposes students to the not unreasonable suggestion that, however well meant, teacher intuition can be wrong.

It is here that we arrive at the second way in which advice to make 'judicious' and 'systematic' use of the L1 is problematic. Where the field has gone beyond a reliance on teacher intuition, and has explicitly described L1-mediated approaches to teaching, there appears to be very little empirical evidence about the substantive effects of these approaches. At least, such empirical evidence that exists tends to stop short of assessing the effects on objective measures of pupil attainment. One can argue about the relative ideological importance of outcomes like test scores or how closely a piece of student writing matches a set of ideals, compared to multilingual pupils' senses of identity or their feelings of inclusion. However, one cannot escape the reality that *both* of these types of outcome are

important to teachers and learners. The asymmetry between observational and experimental research on L1 use in UK classrooms (see the brief discussion on this theme in section 2.10.2) exposes a tendency to ignore the former in favour of the latter.

To get closer to the ideal that all teachers are in a position to provide quality education for their multilingual students, no matter their experience or level of insightfulness, teachers need explicit guidance on what L1-mediated learning looks like in the classroom and clear indications of what happens when these approaches are implemented across different EAL contexts. My thesis addressed directly the problems with the received wisdom about L1-mediated learning as I have defined them here.

The literature review chapter of this work explored the theoretical basis for using learners' L1s in second language learning contexts. It demonstrated that the basis for L1-mediated L2 learning is informed by well-established theories of linguistic interdependence. In turn, these theories, are informed by observational data such as correlations between proficiency in L1 and proficiency in L2, observations about the purposes to which L1 is put when multilingual learners are allowed or instructed to use their L1, and evaluations of bilingual schooling as it compares to monolingual schooling. However, in the orthodoxy that informed the literature review, empirical data about how these theoretical principles and empirical observations can be operationalised to the academic and linguistic advantage of multilingual learners in linguistically diverse primary schools appeared relatively scarce. In essence, the literature review clarified to some extent what systematic, judicious, or principled use of the L1 might entail, but it fell short when asked to provide sufficient empirical evidence to describe the substantive effects of such use. Moreover, much of the evidence that appears to inform the orthodoxy is rooted in markedly different contexts to the linguistically diverse primary schools in which such advice is consumed in the UK. Instead, the main sources of evidence are studies of bilingual schools or EFL classes in national education systems overseas.

Bilingual schools and EFL class contexts tend to be linguistically homogenous when compared with typical primary schools in the UK. Operationalising L1-mediated teaching in those contexts is therefore more straightforward to achieve. In addition, bilingual schools are 'holistically' bilingual. That is, the L1 is afforded a self-contained, systematic curriculum of its own that runs alongside the L2 curriculum. In bilingual schools, L1-mediated strategies are not 'bolted on' to an L2 curriculum. Instead the L1 plays a partner role alongside L2. For these and other potentially important reasons (the socio-linguistic similarity of students, or teachers' proficiency in the L1 of their students, for example), much of this evidence is not applicable to typical UK primary school contexts.

Evidence from evaluations of L1-mediated learning in older learners presents a different kind of uncertainty. We know that people who learn an additional language when they are older are advantaged by cognitive and meta-cognitive skills that are better developed than their comparatively less experienced, younger peers (Lightbown 2008). These better developed skills are attributed in part to the well-established L1 of older learners, something that younger learners do not yet have. Assuming that lessons learned from older learners will apply to younger learners fails to take into account this important developmental difference.

So, while evidence from contexts other than linguistically diverse primary schools in the UK provides an important basis on which to explore the possibility that using the L1 will help in L2 learning in those contexts, it nonetheless leaves important questions unanswered. To begin to evaluate the appropriateness of the advice to teachers, it was necessary to move beyond what was readily available in the literature to assess whether the type of evidence that I have suggested is missing had been overlooked. To do this I prepared a systematic review.

Following an extensive search for relevant literature, only ten studies met the systematic review's inclusion criteria. Of this slim body of evidence, half addressed the use of L1 to support vocabulary learning in the L2. Results of these five studies, on balance, suggested that the approach was beneficial. The remaining five studies meeting the review's inclusion criteria were too dissimilar to allow for any meaningful synthesis, leaving an inconclusive overall picture of the effects of using the L1 as a pedagogical tool for multilingual learners.

The findings of the systematic review represent a huge discrepancy between what is claimed about the use of the L1 and what can be supported by evidence.

Based on the promise suggested by the vocabulary studies included in the systematic review, a randomised crossover trial to compare the use of L1 and English-only to teach English vocabulary to a group of linguistically diverse EAL learners was conducted. It found that, irrespective of the language used to mediate English vocabulary learning, on average new vocabulary was equally well learned. L1 mediation did not confer the advantage often claimed for it.

The implications of the findings of the two primary research elements to this work are that, until such time as the evidence changes, teachers should adopt whatever approach to using the L1s of their EAL learners with which they feel comfortable. If teachers are in a position to operationalise their students' L1s for use in the classroom, this may meet important social justice aims, is unlikely to be harmful, but claims that it improves academic and linguistic outcomes for these students are not well supported by evidence. Conversely, if teachers are not in a position to operationalise the L1s of their students, they should not feel that they are letting them down or wilfully disadvantaging them.



These teachers' time, effort and resources can be directed at teaching strategies for which there is sound empirical evidence.

There is more work to be done on the use of L1 in L2 classrooms, especially linguistically diverse classrooms. In particular, research should be undertaken that allows for analysis of the effects of L1-mediated teaching strategies on different subgroups under the umbrella category of EAL. This would allow for an assessment of whether the bottom-line findings of this study are equally applicable across the population, or whether these types of strategy are particularly effective for some types of learner (for example learners with low levels of English proficiency and high levels of L1 proficiency). In addition, theory should be reviewed to assess the veracity of causal claims regarding the educative effect of using the L1s of multilingual learners in the way so often advocated.

# REFERENCES

Adams' Free Grammar School (1656). 'Statutes, Constitutions and Orders, made and appointed by William Adams, the founder of the Free Grammar School at Newport, in the county of Salop; and at all times to be observed and kept, for the better ordering and government of his said Free Grammar School'. Newport: Adams' Free Grammar School.

Adams, R. and Belam, M. (2016). '#BoycottSchoolCensus: why parents are refusing to reveal their child's nationality.' *The Guardian*, Saturday 8 October. Available at: <https://www.theguardian.com/education/2016/oct/08/boycottschoolcensus-why-parents-are-refusing-to-reveal-their-childs-nationality> [Accessed 3 Jan 2019].

Aldridge, D., Biesta, G., Filippakou, O. and Wainwright, E. (2018). 'Why the nature of educational research should remain contested: A statement from the new editors of the British Educational Research Journal'. *British Educational Research Journal*, 44:1, 1-4.

Alegría de la Colina, A., and del Pilar García Mayo, M. (2009). 'Oral interaction in task-based EFL learning: The use of the L1 as a cognitive tool'. *IRAL - International Review of Applied Linguistics in Language Teaching*, 47:3-4, 325-345.

Altarriba, J. and Mathis, K. M. (1997). 'Conceptual and lexical development in second language acquisition'. *Journal of Memory and Language*, 36:4, 550-568.

Alvarez, J. (1975). *Comparison of academic aspirations and achievement in bilingual versus monolingual classrooms*. Doctoral dissertation of Texas, Austin.

Ashcroft, R. (2000). 'Giving medicine a fair trial: Trials should not second guess what patients want'. *BMJ: British Medical Journal*, 320:7251, 1686.

Associated Press (2009). 'Kurdish spoken in challenge to Turkey Politician violates law that bars the language in official places'. *NBCNews.com*. Available at: [http://www.nbcnews.com/id/29371366/ns/world\\_news-europe/t/kurdish-spoken-challenge-turkey/#.Us6-EHl2nSw](http://www.nbcnews.com/id/29371366/ns/world_news-europe/t/kurdish-spoken-challenge-turkey/#.Us6-EHl2nSw) [Accessed 9 Jan 2014].

Atkinson, D. (1987). 'The mother tongue in the classroom: a neglected resource?' *ELT Journal*, 41:4, 241-247.

Atkinson, D. (1993). 'Teaching in the target language: a problem in the current orthodoxy'. *The Language Learning Journal*, 8, 2-5.

- Atwill, K., Blanchard, J., Gorin, J. S., and Burstein, K. (2007). 'Receptive Vocabulary and Cross-Language Transfer of Phonemic Awareness in Kindergarten Children'. *Journal of Educational Research*, 100:6, 336–345.
- Auerbach, E. R. (1993). 'Reexamining English Only in the ESL Classroom'. *TESOL Quarterly*, 27:1, 9 - 32.
- Azari, F., Abdullah, F.S., Heng, C.W., and Hoon, T.B. (2012). 'Effects of Glosses on Vocabulary Gain and Retention among Tertiary Level EFL Learners'. Available at: <http://files.eric.ed.gov/fulltext/ED533228.pdf> [Accessed 30 April 2018].
- Bailey, N., Madden, C., and Krashen, S. D. (1974). 'Is there a "natural sequence" in adult second language learning?' *Language Learning*, 24:2, 235-243.
- Basson, M. and le Cordeur, M. (2013). 'Die bevordering van woordeskat en leesbegrip by Xhosa-moedertaalsprekers in graad 4-6 in Afrikaanse skole'. *Tydskrif vir Geesteswetenskappe*, 53:3, 377 - 390.
- Basson, M. and le Cordeur, M. (2014). 'Effective Support to Xhosa Mother Tongue Speakers in Afrikaans Classes'. *Per Linguam*, 30:1, 109–127.
- Baumann, J.F. and Graves, M.F. (2010). 'What is academic vocabulary?' *Journal of Adolescent and Adult Literacy*, 54:1, 4-12.
- BBC (2014). 'KS2 Bitesize'. Available at: <http://www.bbc.co.uk/bitesize/ks2/> [Accessed 4 July 2018].
- Bell Foundation (2017). 'EAL assessment framework for schools: research and development'. Available at: <https://www.bell-foundation.org.uk/wp-content/uploads/2018/07/EAL-Assessment-Framework-Research-and-Development.pdf>. [Accessed 17 Dec 2018].
- Benson, C. (2002). 'Transfer/Cross-linguistic influence'. *ELT Journal*, 56:1, 68-70.
- BERA (2018). 'Ethical Guidelines for Educational Research, fourth edition'. Available at: <https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018> [Accessed 20 July 2018].
- Berman, R. (1994). 'EFL essay writing: L1 Versus L2 Instruction'. Presented at the 28th Annual Meeting of the Teachers of English to Speakers of Other Languages. March 8-12, Baltimore, MD.
- Bialystok, E., Luk, G., and Kwan, E. (2005). 'Bilingualism, biliteracy, and learning to read: Interactions among languages and writing systems'. *Scientific Studies of Reading*, 9, 43-61.

Bialystok, E., McBride-Chang, C., and Luk, G. (2005). 'Bilingualism, language proficiency, and learning to read in two writing systems'. *Journal of Educational Psychology*, 97, 580-590.

Bilingualism: Language and Cognition (2018). 'Instructions for contributors'. Available at: <https://www.cambridge.org/core/journals/bilingualism-language-and-cognition/information/instructions-contributors> [Accessed 13 September 2018].

Blumenfeld, H. K., and Marian, V. (2013). 'Parallel language activation and cognitive control during spoken word recognition in bilinguals'. *Journal of Cognitive Psychology*, 25:5, 547-567.

Bourne, J. (2002). 'Home languages in the literacy hour in the national literacy strategy: Supporting pupils learning English as an additional language'. London: DfES.

Bowers, P. N., Kirby, J. R., and Deacon, S. H. (2010). 'The Effects of Morphological Instruction on Literacy Skills'. *Review of Educational Research*, 80(1), 144–179.

Bozorgian, H., and Pillay, H. (2013). 'Enhancing Foreign Language Learning through Listening Strategies Delivered in L1: An Experimental Study'. *International Journal of Instruction*, 6:1, 105–122.

Brunton, G., Stansfield, C. and Thomas, J. (2012). 'Finding Relevant Studies'. In Gough, D., Oliver, S. and Thomas, J. (eds.) *An Introduction to Systematic Reviews*. London: SAGE, pp. 107-134

Burgoyne, K., Whiteley, H.E. and Hutchinson, J.M. (2011). 'The development of comprehension and reading-related skills in children learning English as an additional language and their monolingual English-speaking peers'. *British Journal of Educational Psychology*, 81, 344-354.

Bush, G.W. (2000). 'Speech delivered at the NAACP's 91st annual convention, July 10, 2000'. Transcript available at: <http://www.washingtonpost.com/wp-srv/onpolitics/elections/bushtext071000.htm> [Accessed 19 Apr 2018].

Butzkamm, W. (2003). 'We only learn language once. The role of the mother tongue in FL classrooms: death of a dogma'. *Language Learning Journal*, 28:1, 29-39.

Butzkamm, W. (2007). 'Native Language Skills as a Foundation for Foreign Language Learning'. In Kindermann, W. (ed.), *Transcending boundaries. Essays in honour of Gisela Hermann-Brennecke*. Berlin: Lit Verlag, pp. 71–85.

Camó, A.C. and Ballester, E.P. (2015). 'The Effects of Using L1 Translation on Young Learners' Foreign Language Vocabulary Learning'. *Elia*, 15, 109-134.

Campbell Collaboration (2014). 'Methodological Expectations of Campbell Collaboration Intervention Reviews'. Available at: <https://campbellcollaboration.org/library/campbell-methods-reporting-standards.html> [Accessed 8 Feb 2019].

Campbell Collaboration (2017). 'Campbell Collaboration Systematic Reviews: Policies and Guidelines'. Available at: <https://www.campbellcollaboration.org/library/campbell-collaboration-systematic-reviews-policies-and-guidelines.html> [Accessed 27 Aug 2018].

Campbell, D.T. and Stanley, J.C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally & Co.

Campeau, P. L., Roberts, A., Oscar, H., Bowers, J. E., Austin, M., and Roberts, S. J. (1975). *The identification and description of exemplary bilingual education programs*. Palo Alto, CA: American Institutes for Research.

Carlisle, J.F. (1995). 'Morphological awareness and early reading achievement'. In Feldman, L.B. (ed.), *Morphological aspects of language processing*. Hillsdale, NJ: Lawrence Erlbaum, pp. 189–210.

Carlisle, J.F. (2003). 'Morphology matters in learning to read: A commentary'. *Reading Psychology*, 24, 291–332.

Carroll, J.B. (1962). *The prediction of success in intensive foreign language training*. Pittsburgh: University of Pittsburgh Press.

Castles, A., Rastle, K. and Nation, K. (2018). 'Ending the Reading Wars: Reading Acquisition From Novice to Expert'. *Psychological Science in the Public Interest*, 19:1, 5-51.

Catalogue of Bias Collaboration, Nunan, D., Bankhead, C., Aronson, J.K. (2017a). *Selection bias. Catalogue Of Bias*. Available at: <http://www.catalogofbias.org/biases/selection-bias/> [Accessed 30 Dec 2018].

Catalogue of Bias Collaboration, Nunan, D., Bankhead, C., Aronson, J.K. (2017b). *Volunteer bias. Catalogue Of Bias*. Available at: <http://www.catalogofbias.org/biases/volunteer-bias> [Accessed 30 Dec 2018].

Celic, C. and Seltzer, K. (2013). 'Translanguaging: A CUNY-NYSIEB Guide for Educators'. Available at: <https://www.cuny-nysieb.org/wp-content/uploads/2016/04/Translanguaging-Guide-March-2013.pdf> [Accessed 16 Aug 2018].

Celic, C. and Seltzer, K. (2012). *Translanguaging: A CUNY-NYSIEB Guide for Educators*. New York: CUNY-NYSIEB.

Centre for Open Science (no date). *Show Your Work. Share Your Work. Advance Science. That's OPEN SCIENCE*. Available at: <http://cos.io> [Accessed 16 Feb 2019].

CESE (2017). 'Cognitive load theory: Research that teachers really need to understand. Sydney: Centre for Education Statistics and Evaluation'. Available at: <https://www.cese.nsw.gov.au/publications-filter/cognitive-load-theory-research-that-teachers-really-need-to-understand> [Accessed 8 May 2018].

Chalmers, H (2014). 'Harnessing linguistic diversity in polylingual British curriculum schools. Do L1 mediated home learning tasks improve learning outcomes for bilingual children? A randomised trial'. Masters Dissertation. Oxford Brookes University. Available at: [https://englishagenda.britishcouncil.org/sites/default/files/filefield\\_paths/harnessing\\_linguistic\\_diversity\\_for\\_bc\\_v2.pdf](https://englishagenda.britishcouncil.org/sites/default/files/filefield_paths/harnessing_linguistic_diversity_for_bc_v2.pdf) [Accessed 8 Feb 2019].

Chalmers, I., Bracken, M.B., Djulbegovic, B., Garattini, S., Grant, J., Metin Gülmezoglu, A.M., Howells, D.W., Ioannidis, J.P.A., and Oliver, S. (2014). 'How to increase value and reduce waste when research priorities are set'. *The Lancet*, 383:9912, 156-165.

Chalmers, I. and Lindley, R. (2001). 'Double standards on informed consent to treatment'. In Doyal, L., and Tobias, J.S., (eds.), *Informed Consent in Medical Research*. London: BMJ Publications, pp. 266-275.

Cheng, Y.-H. and Good, R. (2009). 'L1 glosses: Effects on EFL learners' reading comprehension and vocabulary retention'. *Reading in a Foreign Language*, 21:2, 119-142.

Chuang, H.-K., Joshi, R. M., and Dixon, L. Q. (2011). 'Cross-Language Transfer of Reading Ability'. *Journal of Literacy Research*, 44:1, 97-119.

Chumak-Horbatsch, R. (2012). *Linguistically Appropriate Practice: A Guide for Working with Young Immigrant Children*. Toronto: University of Toronto Press.

Cochrane (2016). *About Us*. Available at: <http://www.cochrane.org/about-us>. [Accessed 5 Jun 2016].

Coe, R. (2014). *Would you let this test into your classroom?* Available at: <http://www.cem.org/blog/would-you-let-this-test-into-your-classroom/> [Accessed 18 Dec 2018].

- Cohen, A. D. (1975). *A sociolinguistic approach to bilingual education*. Rowley, MA: Newbury House Press.
- Cohen, L., Manion, L., and Morrison, K. (2007). *Research Methods in Education (sixth edition)*. Abingdon: Routledge.
- Collier, V., and Thomas, W. P. (2017). 'Validating the power of bilingual schooling: Thirty-two years of large-scale, longitudinal research'. *Annual Review of Applied Linguistics*, 3, 1–15.
- Colombo, M. and Furbush, D. (2009). *Teaching English language learners: content and language in middle and secondary mainstream*. Thousand Oaks, CA: SAGE.
- Comeau, L., Cormier, P., Grandmaison, É., and Lacroix, D. (1999). 'A longitudinal study of phonological processing skills in children learning to read in a second language'. *Journal of Educational Psychology*, 91, 29–43.
- CRE (1986). *Teaching English as a Second Language: Report of a Formal Investigation in Calderdale Local Education Authority*. London: Commission for Racial Equality.
- Connolly, P. (2015). 'The trials of evidence-based practice in education'. *Research Intelligence*. Conference Special Edition, October, 6.
- Connolly, P., Biggart, A., Miller, S., O'Hare, L., and Thurston, A. (2017). *Using randomised controlled trials in education*. London: SAGE.
- Connolly, P., Keenan, C., and Urbanska, K. (2018). 'The trials of evidence-based practice in education: a systematic review of randomised controlled trials in education research 1980–2016'. *Educational Research*, x:x 1-16.
- CONSORT (no date). *Welcome to the CONSORT Website*. Available at: <http://www.consort-statement.org> [Accessed 3 Dec 2018].
- Conteh, J. (2015). *The EAL teaching book (Second edition)*. Los Angeles: SAGE.
- Cook, V. (1999). 'Going beyond the native speaker in language teaching'. *TESOL Quarterly*, 33:2, 185-209.
- Cook, V. (2001a). 'Linguistics and Second Language Acquisition: One Person with Two Languages'. In Aronoff, M. and Rees-Miller, J. (eds.), *The Blackwell Handbook of Linguistics*, Oxford: Blackwell pp. 488-511.

- Cook, V. (2001b). 'Using the First Language in the Classroom'. *Canadian Modern Language Review*, 57:3, 402–423.
- Costley, T. (2013). 'English as an additional language, policy and the teaching and learning of English in England'. *Language and Education*, 28:3, 276-292.
- Covey, D.D. (1973). 'An analytical study of secondary freshmen bilingual education and its effects on academic achievement and attitudes of Mexican American students'. Doctoral dissertation, Arizona State University, Tempe. Available at <https://eric.ed.gov/?id=ED079671> [Accessed 8 Feb 2019].
- Crozier, G. and Davies, J. (2007). 'Hard to reach parents or hard to reach schools? A discussion of home—school relations, with particular reference to Bangladeshi and Pakistani parents'. *British Educational Research Journal*, 33:3, 295-313.
- Cummins, J. (1979). 'Linguistic interdependence and the educational development of bilingual children'. *Review of Educational Research*, 49:2, 222-251.
- Cummins, J. (1980). 'The construct of language proficiency in bilingual education'. In Alatis, J.E. (ed.) *Georgetown University Round Table on Languages and Linguistics 1970*. Washington, D.C.: Georgetown University Press, pp. 81-103.
- Cummins, J. (1981). 'The role of primary language development in promoting educational success for language minority students'. In California State Department of Education (ed.), *Schooling and language minority students: A theoretical framework*, pp. 3-49.
- Cummins, J. (1998). 'Immersion education for the millennium: What have we learned from 30 years of research on second language immersion?' In Childs, M.R. and Bostwick, R.M. (eds.), *Learning through two languages: Research and practice. Second Katoh Gakuen international symposium on immersion and bilingual education*. Numazu, Japan: Katoh Gakuen, pp. 34–47.
- Cummins, J. (2001). 'Bilingual children's mother tongue: Why is it important for education?' *Sprogforum* 7:19, 15-20.
- Cummins, J. (2007). 'Rethinking monolingual instructional strategies in multilingual classrooms'. *Canadian Journal of Applied Linguistics*, 10:2, 221-240.
- Cummins, J. (2011). 'The intersection of cognitive and sociocultural factors in the development of reading comprehension among immigrant students'. *Reading and Writing*, 25:8, 1973–1990.



Cummins, J. (2016). 'Teaching for Transfer in Multilingual School Contexts'. In García, O., Lin, A., and May, S. (eds.), *Encyclopedia of Language and Education, 3<sup>rd</sup> Edition, Volume 5: Bilingual Education*. New York: Springer, pp. 103–115.

Cummins, J., and Mulcahy, R. (2016). 'Orientation to Language in Ukrainian-English Bilingual Children'. *Child Development*, 49:4, 1239–1242.

Darcy, N.T. (1953). 'A Review of the Literature on the Effects of Bilingualism Upon the Measurement of Intelligence'. *The Journal of Genetic Psychology*, 82:1, 21–57.

De Bot, K., Lowie, W., and Verspoor, M. (2005). *Second Language Acquisition: an advanced resource book*. Abingdon: Routledge.

de Bruin, A., Treccani, B. and Della Samma, S. (2015). 'Cognitive advantage in bilingualism. An example of publication bias?' *Psychological Science*, 26:1, 99-107.

de la Campa, J. and Nassaji, H. (2009). 'The amount, purpose, and reasons for using L1 in L2 classrooms'. *Foreign Language Annals*, 42:4, 742-759.

Deacon, S.H., Wade-Woolley, L., and Kirby, J. (2007). 'Crossover: The role of morphological awareness in French immersion children's reading.' *Developmental Psychology*, 43, 732–746.

Deaton, A. and Cartwright, N. (2017).

Demie, F. (2018). 'English language proficiency and attainment of EAL (English as second language) pupils in England'. *Journal of Multilingual and Multicultural Development*. 0:0, 1-13.

Demie, F., McDonald, J., and Hau, A. (2016). *Language Diversity and Attainment in Secondary Schools in England*. London: Lambeth Research and Statistics Unit. Available at: [https://www.lambeth.gov.uk/rsu/sites/www.lambeth.gov.uk/rsu/files/Language\\_%20Diversity\\_and\\_Attainment\\_in\\_Secondary\\_Schools\\_2014.pdf](https://www.lambeth.gov.uk/rsu/sites/www.lambeth.gov.uk/rsu/files/Language_%20Diversity_and_Attainment_in_Secondary_Schools_2014.pdf) [Accessed 6 Feb 2018].

Dennis, L.R., Kratch, S.K., McCreery, M.P., and Navarro, S. (2018). 'The Student Oral Language Observation Matrix: A Psychometric Study With Preschoolers'. *Assessment for Effective Intervention*. 00:0, 1-8.

Department for Children Schools and Families (2009). *Deprivation and Education The evidence on pupils in England, Foundation Stage to Key Stage 4*. Available at: <https://dera.ioe.ac.uk/9431/1/DCSF-RTP-09-01.pdf> [Accessed 30 Dec 2018].

DfE (2013). *The national curriculum in England Key stages 1 and 2 framework document*. Available at:

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/425601/PRIMARY\\_national\\_curriculum.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/425601/PRIMARY_national_curriculum.pdf) [Accessed 10 Oct 2017].

DfE (2014). *School census spring and summer 2014 guide for all-through schools*. Available at:

[http://dera.ioe.ac.uk/19515/1/2014\\_Spring\\_and\\_Summer\\_School\\_Census\\_Guide\\_for\\_all-through\\_\\_Version1.1.pdf](http://dera.ioe.ac.uk/19515/1/2014_Spring_and_Summer_School_Census_Guide_for_all-through__Version1.1.pdf) [Accessed 31 Jul 2017].

DfE (2016a). *School census 2016 to 2017 Guide, Version 1.4 November 2016*. Available at:

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/564384/School\\_census\\_guide\\_2016\\_2017\\_v1\\_4.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/564384/School_census_guide_2016_2017_v1_4.pdf) [Accessed 22 Nov 2016].

DfE (2016b). *Key stage 2 tests: 2016 English grammar, punctuation and spelling test materials*.

Available at: <https://www.gov.uk/government/publications/key-stage-2-tests-2016-english-grammar-punctuation-and-spelling-test-materials>. [Accessed 4 Jul 2018].

DfE (2017a). *Schools, pupils and their characteristics: January 2017 (SFR 28/2017)*. London:

Department for Education. Available at:

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/623124/SFR28\\_2017\\_Main\\_Text.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/623124/SFR28_2017_Main_Text.pdf) [Accessed 17 Jul 2017].

DfE (2017b) *Pupil nationality, country of birth and proficiency in English. Summary Report*. Available

at:

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/665127/Data\\_on\\_pupil\\_nationality\\_\\_country\\_of\\_birth\\_and\\_proficiency.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/665127/Data_on_pupil_nationality__country_of_birth_and_proficiency.pdf) [Accessed 6 Feb 2018].

DfE (2018a). *Schools, pupils and their characteristics: January 2018*. Available at:

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/719226/Schools\\_Pupils\\_and\\_their\\_Characteristics\\_2018\\_Main\\_Text.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/719226/Schools_Pupils_and_their_Characteristics_2018_Main_Text.pdf). [Accessed 24 Nov 2018].

DfE (2018b). *Reducing teacher workload*. Available at:

<https://www.gov.uk/government/publications/reducing-teachers-workload/reducing-teachers-workload>. [Accessed 1 Jan 2019].

Doebler, L.K. and Mardis, L.J. (1980-1981). 'Effects of a bilingual education program for Native American children'. *NABE Journal*, 5:2, 23-28.

Draper, S. (2018). *Effect size*. Available at: <http://www.psy.gla.ac.uk/~steve/best/effect.html> [Accessed 27 Aug 2018].

Duarte, J. (2016). 'Translanguaging in mainstream education: a sociocultural approach'. *International Journal of Bilingual Education and Bilingualism*, 0:0, 1-15.

Dunston, P.J. and Tyminski, A.M. (2013). 'What's the big deal about vocabulary?' *Mathematics Teaching in the Middle School*, 19:1, 38-45.

Echevarría, J., Vogt, M., and Short, D.J. (2013). *Making content comprehensible for English learners The SIOP® Model 4th Edition*. Boston: Pearson.

Edstrom, A. (2006). 'L1 Use in the L2 Classroom: One Teacher's Self-Evaluation'. *Canadian Modern Language Review*, 63:2, 275-292.

EEF (2012). *Annual Report 2011/12*. Available at: [https://educationendowmentfoundation.org.uk/public/files/Annual\\_Reports/EEF\\_Annual\\_Report\\_2011-12\\_-\\_interactive.pdf](https://educationendowmentfoundation.org.uk/public/files/Annual_Reports/EEF_Annual_Report_2011-12_-_interactive.pdf). [Accessed 1 Jan 2019].

EEF (2018a). *What Works Network*. Available at: <https://educationendowmentfoundation.org.uk/about/what-works-network/#closeSignup> [Accessed 30 Jul 2018].

EEF (2018b). *For Grantees*. Available at: <https://educationendowmentfoundation.org.uk/pdf/generate/?u=https://educationendowmentfoundation.org.uk/pdf/content/?id=314andc=314andt=Our%20Approach%20to%20Evaluationands=andmode=embed> [Accessed 31 May 2018].

Elliot, J. (2001). 'Making evidence-based practice educational'. *British Educational Research Journal*, 27:5, 555-574.

Ouzzani, M., Hammady, H., Fedorowicz, Z., and Elmagarmid, A. (2016). 'Rayyan — a web and mobile app for systematic reviews'. *Systematic Reviews*, 5:210, 1-10.

Encyclopædia Britannica, Inc. (2018). *Britannica Kids*. Available at: <https://kids.britannica.com> [Accessed 4 Jul 2018].

Epstein, N. (1977). *Language, ethnicity, and the schools: Policy alternatives for bilingual-bicultural education*. Washington, DC: Institute for Educational Leadership, George Washington University.

Espinosa, C., Ascenzi-Moreno, L., and Vogel, S. (2016). *A Translanguaging Pedagogy for Writing: A CUNY-NYSIEB Guide for Educators*. Available at: <https://www.cuny-nysieb.org/wp-content/uploads/2016/05/TLG-Pedagogy-Writing-04-15-16.pdf> [Accessed 16 Aug 2018].

ESRC (2015). *ESRC Framework for research ethics*. Available at: <http://www.esrc.ac.uk/files/funding/guidance-for-applicants/esrc-framework-for-research-ethics-2015/> [Accessed 16 Jan 2017].

Evans, I., Thornton, H., Chalmers, I. and Glasziou, P. (2011). *Testing Treatments, Better research for better healthcare. Second Edition*. London: Pinter and Martin.

Evans, M., Schneider, C., Arnot, M., Fisher, L., Forbes, K., Mei, H. and Liu, Y. (2016). *Language development and school achievement: Opportunities and challenges in the education of EAL students (Executive Summary)*. Cambridge: The Bell Foundation.

Faculty of Public Health (2010). *What is public health?* Available at: [http://www.fph.org.uk/what\\_is\\_public\\_health](http://www.fph.org.uk/what_is_public_health) [Accessed 22 Jul 2016].

Feiler, A. (2010). *Engaging 'hard to reach' parents Teacher-parent collaboration to promote children's learning*. Chichester: Wiley-Blackwell.

Fernandez, C.V., Kodish, E. and Weijer, C. (2003). 'Informing Study Participants of Research Results: An Ethical Imperative'. *IRB: Ethics and Human Research*, 25:3, 12-19.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. SAGE: London.

Flesch, R. (1979). *How to write plain English*. New York: Harper Collins.

Flynn, N. and Curdt-Christiansen, X.-L. (2018). 'Intentions versus enactment: making sense of policy and practice for teaching English as an additional language.' *Language and Education*, 32:5, 410-427.

Freyer, D.A., Frederick, W.C. and Klausmeier, H.J. (1969). *A Schema for Testing the Level of Concept Mastery. Technical Report No. 16*. Madison: University of Wisconsin Research and Development Center for Cognitive Learning.

Friedenberg, J.E. (1984). 'The Effects of Simultaneous Bilingual Reading Instruction on the Development of English Reading Skills'. *Acta Paedologica*, 1:2, 1-4.

Fung, I.Y.Y., Wilkinson, I., and Moore, D. (2003). 'L1-assisted reciprocal teaching to improve ESL students' comprehension of English expository text'. *Learning and Instruction*, 13:1, 1-31.

- Furedi, F. (2013). *Keep the scourge of scientism out of schools*. Spiked, 9 September. Available at: <http://www.spiked-online.com/newsite/article/14005> [Accessed 30 Jul 2018].
- Gagnier J.J., Kienle, G., Altman, D.G., Moher, D., Sox, H., Riley, D.; the CARE Group (2013). 'The CARE Guidelines: Consensus-based Clinical Case Reporting Guideline Development'. *Journal of Clinical Epidemiology*, 67:1, 46-51.
- Gallagher, E. (2008). *Equal Rights to the Curriculum: Many Languages, One Message*. Clevedon: Multilingual Matters.
- García, O. and Li Wei (2014). *Translanguaging: Language, Bilingualism and Education*. Basingstoke: Palgrave Macmillan.
- García, O., and Sánchez, M.T. (2015). 'Transforming schools with emergent bilinguals: The CUNY-NYSIEB Project'. In Dirim, I., Gogolin, I., Knorr, D., Krüger-Potratz, M., Lengyel, D., Reich, H., and Weiße, W. (eds.), *Impulse für die Migrationsgesellschaft: Bildung, Politik und Religion*. Münster: Waxman, pp. 80–94.
- Gayle, D. (2016). 'Pupil data shared with Home Office to 'create hostile environment' for illegal migrants'. *The Guardian*, Thursday 15 December. Available at: <https://www.theguardian.com/uk-news/2016/dec/15/pupil-data-shared-with-home-office-to-identify-illegal-migrants> [Accessed 20 Jul 2018].
- Gebauer, S.K., Zaunbauer, A.C.M., and Möller, J. (2013). 'Cross-language transfer in English immersion programs in Germany: Reading comprehension and reading fluency'. *Contemporary Educational Psychology*, 38:1, 64–74.
- Gersten, R., and Woodward, J. (1995). 'A longitudinal study of transitional and immersion bilingual education programs in one district'. *Elementary School Journal*, 95:3, 223-239.
- Gibb, N. (2016). 'Pupils: Personal Records'. Parliament: written questions and written answers, 25 July, 42942. Available at: <https://www.parliament.uk/business/publications/written-questions-answers-statements/written-question/Commons/2016-07-15/42942/> [Accessed 20 Jul 2018].
- Gibbons, P. (2009). *Multilingual learners literacy and thinking*. Portsmouth, NH: Heinemann.
- Gill, D. (2013). 'Practitioners' and parents' perceptions and attitudes about bilingual education'. *Race Equality Teaching*, 31:3, 19-24.

Glass, G.V. (1976). 'Primary, secondary, and meta-analysis of research'. *Educational Researcher*, 5:10 3–8.

Goldacre, B. (2013). *Building Evidence into Education*. Available at: <https://www.gov.uk/government/news/building-evidence-into-education> [Accessed 30 Jul 2018].

Goldacre, B., Drysdale, H., Powell-Smith, A., Dale, A., Milosevic, I., Slade, E., Hartley, P., Marston, C., Mahtani, K., and Heneghan, C. (2019). *Methods*. Available at: <http://compare-trials.org/methods#whyitmatters> [Accessed 16 Feb 2019].

Goldenberg, C. (2008). 'Teaching English Language Learners What the research does - and does not – say'. *American Educator*, 32:2, 8-23 and 42-44.

Gorard, S. (2013). *Research Design, Creating robust approaches for the social sciences*. London: SAGE.

Gorard, S. (2014) 'A proposal for judging the trustworthiness of research findings'. *Radical Statistics*, 110, 47–59.

Gore, M.C. (2009). *Inclusion strategies for secondary classrooms Keys for struggling learners*. Thousand Oaks, CA: Corwin.

Gough, D., Oliver, S. and Thomas, J. (2012). *An Introduction to Systematic Reviews*. London: SAGE.

GOV.UK (2014). *Elizabeth Truss speaks about improving teaching*. Transcript available at: <https://www.gov.uk/government/speeches/elizabeth-truss-speaks-about-improving-teaching> [Accessed 18 Dec 2018].

Gov.UK (no date). *Ofsted*. Available at: <https://www.gov.uk/government/organisations/ofsted/about> [Accessed 5 Dec 2016].

Graves, M.F. and Penn, M.C., (1986). 'Costs and Benefits of Various Methods of Teaching Vocabulary'. *Journal of Reading*, 29:7, 596-602.

Grosjean, F. (2010). *Bilingual: Life and Reality*. Cambridge, MA: Harvard University Press.

Gueron, J.M. (2008). 'The politics of random assignment: implementing studies and impacting policy'. *Journal of Children's Services*, 3:1, 14-26.

Hall, G. and Cook, G. (2012). 'Own-language use in language teaching and learning'. *Language Teaching*, 45:03, 271–308.

- Hammerly, H. (1991). *Fluency and accuracy. Toward balance in language teaching and learning*. Clevedon: Multilingual Matters.
- Harbord, J. (1992). 'The use of the mother tongue in the classroom'. *ELT Journal*, 46:4, 350-355.
- Harris, J. (2018). 'Ukip may have collapsed, but where it led others will follow'. *The Guardian*, Monday 23 April. Available at: <https://www.theguardian.com/commentisfree/2018/apr/23/ukip-collapsed-online-activism-immigration> [Accessed 17 Dec 2018].
- Haynes, J. and Zacarian, D. (2010). *Teaching English language learners across the content areas*. Alexandria, VA: ASCD.
- He, A.E. (2012). 'Systematic use of mother tongue as learning/teaching resources in target language instruction'. *Multilingual Education*, 2:1, 1-15.
- Hesson, S., Seltzer, K., and Woodley, H.H. (2014). *Translanguaging in Curriculum and Instruction: A CUNY-NYSIEB Guide for Educators*. Available at: <https://www.cuny-nysieb.org/wp-content/uploads/2016/05/Translanguaging-Guide-Curr-Inst-Final-December-2014.pdf> [Accessed 16 Aug 2018].
- Hillner, B. (2005). "*Bad Policy and Bad Law*": *The Shortcomings of the No Child Left Behind Act in Bilingual Educational Policy and its Frustration of the Equal Protection Clause*. Available at: <http://www.educationlawconsortium.org/forum/2005/papers/hillner.pdf> [Accessed 9 Apr 2018].
- Hitchens, C. (2003). 'Mommie Dearest'. *Slate*, Monday 20 October. Available at: [http://www.slate.com/articles/news\\_and\\_politics/fighting\\_words/2003/10/mommie\\_dearest.html](http://www.slate.com/articles/news_and_politics/fighting_words/2003/10/mommie_dearest.html) [Accessed 10 May 2016].
- Howatt, A. (1984). *A History of English Language Teaching*. Oxford: Oxford University Press.
- Huennekens, M.E. (2013). *The Cross-Linguistic Effects of Dialogic Reading on Young English Language Learners*. Unpublished PhD Thesis. Virginia Commonwealth University, Virginia, USA.
- Husain, F. (2018). *Family Skills Evaluation report and executive summary*. Available at: [https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation\\_Reports/Family\\_Skills.pdf](https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Family_Skills.pdf) [Accessed 17 Dec 2018].

Huzar, H. (1973). *The effects of an English-Spanish primary grade reading program on second and third grade students*. Unpublished master's thesis, Rutgers University, Camden, NJ. Available at: <http://files.eric.ed.gov/fulltext/ED085683.pdf> [Accessed 26 Oct 2017].

Immigration Act 2014, c. 22. Available at: [http://www.legislation.gov.uk/ukpga/2014/22/pdfs/ukpga\\_20140022\\_en.pdf](http://www.legislation.gov.uk/ukpga/2014/22/pdfs/ukpga_20140022_en.pdf) [Accessed 18 Dec 2018].

Institute of Education Sciences (no date). *About the WWC*. Available at: <http://ies.ed.gov/ncee/wwc/aboutus.aspx> [Accessed 31 May 2016].

Ipsos/MORI (2016). *Concern about immigration rises as EU vote approaches*. Available at: <https://www.ipsos.com/ipsos-mori/en-uk/concern-about-immigration-rises-eu-vote-approaches> [Accessed 17 Dec 2018].

Jackson, D.B. (2015). 'A Targeted Role for L1 in L2 Vocabulary Acquisition with Mobile Learning Technology'. *TESOL Arabia Perspectives*, 23:1, 6-11.

Jacobs, J. (1998). 'Latino vote is key to depth of 227's victory'. *San José Mercury News*, Thursday 14 May. Available at: <http://www.onenation.org/opinion/latino-vote-is-key-to-depth-of-227-s-victory/> [Accessed 2 May 2016].

Jacobs, G.M., Dufon, P., and Hong, F.C. (1994). 'L1 and L2 vocabulary glosses in L2 reading passages: their effectiveness for increasing comprehension and vocabulary knowledge'. *Journal of Research in Reading*, 17:1, 19-28.

Jenkins, J.R., Fuchs, L.S., van den Broek, P., Espin, C., and Deno, S.L. (2003). 'Sources of individual differences in reading comprehension and reading fluency'. *Journal of Educational Psychology*, 95, 719–729.

Kajonkiet International School (2014). *English Speaking Policy Early Years Foundation Stage and Primary*. Available at: <http://www.kisphuket.ac.th/wp-content/uploads/2015/08/English-Speaking-Policy-6.pdf> [Accessed 2 May 2016].

Kale, J. and Luke, A. (1991). 'Doing things with words in early language socialisation'. In Furniss, E. and Green, P. (eds) *The Literacy Agenda: issues for the nineties*. Armadale, Victoria: Eleanor Curtain Publishing.

Kaufman, M. (1968). 'Will instruction in reading Spanish affect ability in reading English?' *Journal of Reading*, 11, 521-527.



- Kenner, C., Al-Azami, S., Gregory, E., and Ruby, M. (2008). 'Bilingual poetry: expanding the cognitive and cultural dimensions of children's learning'. *Literacy*, 42:2, 92-100.
- Keung, Y.-C., and Ho, C. S.-H. (2009). 'Transfer of reading-related cognitive skills in learning to read Chinese (L1) and English (L2) among Chinese elementary school children'. *Contemporary Educational Psychology*, 34:2, 103–112.
- Kibler, A. (2010). 'Writing through two languages: First language expertise in a language minority classroom'. *Journal of Second Language Writing*, 19:3, 121–142.
- Kim, H.-Y. (2013). 'Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis'. *Restorative dentistry & endodontics*. 38:1, 52–54.
- Kincaid, J.P., Fishburne, R.P., Rogers, R.L., and Chissom, B.S. (1975). 'Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. Research Branch Report 8–75'. Chief of Naval Technical Training: Naval Air Station Memphis.
- Knight, P. (2001). 'The development of EFL methodology'. In Candlin, C.N. and Mercer, N. (eds.), *English Language Teaching in its Social Context A Reader*. London: Routledge, pp. 147-166.
- Kobayashi, M. (2003). 'The role of peer support in ESL students' accomplishment of oral academic tasks'. *The Canadian Modern Language Review*, 59:3, 337-368.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press Inc.
- Krashen, S. and Terrell, T. (1983). 'The Natural Approach Language Acquisition in the Classroom'. New York: Phoenix ELT.
- Krashen, S., and McField, G. (2005). 'What Works? Reviewing the Latest Evidence on Bilingual Education'. *Language Learner*, 1:2, 7-10.
- Kreidler, C.W. (2014). *Introducing English Semantics, Second Edition*. London and New York: Routledge.
- Kroll, J.F., and Bialystok, E. (2013). 'Understanding the consequences of bilingualism for language processing and cognition'. *Journal of Cognitive Psychology*, 25, 497-514.

- Kuska, S.K., Zaunbauer, A.C.M., and Möller, J. (2010). 'Sind Immersionschüler wirklich leistungsstärker? Ein Lernexperiment [Do immersion students really perform better? A learning experiment]'. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 42, 143-153.
- Lado, R. (1957). *Linguistics across cultures: applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.
- Lambert, W.E., and Tucker G.R. (1972). *Bilingual Education of Children: The St Lambert Experiment*. Michigan: Newbury House.
- Lambert, W.E., and Macnamara, J. (1969). 'Some cognitive consequences of following a first-grade curriculum in a second language'. *Journal of Educational Psychology*, 60:2, 86–96.
- Lambert, W.E., Tucker, G.R., and d'Anglejan, A. (1973). 'Cognitive and Attitudinal Consequences of Bilingual Schooling: The St. Lambert Project Through Grade Five'. *Journal of Educational Psychology*, 66:2, 141–159.
- Lee, J.H. (2012). 'Implications for language diversity in instruction in the context of target language classrooms: Development of a preliminary model of the effectiveness of teacher code-switching'. *English Teaching: Practice and Critique*, 11:4, 137-160.
- Lee, J.H. and Macaro, E. (2013). 'Investigating age in the use of L1 or English-only instruction: Vocabulary acquisition by Korean EFL learners'. *The Modern Language Journal*, 97:4, 887–901.
- Legarreta-Marcaida, D. (1981). 'Effective use of primary language in the classroom'. In California State Department of Education (eds.), *Schooling and language minority students: A theoretical framework*, pp. 83-116.
- Lewis, J., and Reader, T. (2009). 'How can we help adolescent readers meet the challenges of academic text?' In Lewis, J. (ed.), *Essential questions in adolescent literacy: Teachers and researchers describe what works in classrooms*, New York: The Guilford Press, pp. 101–131.
- Li Wei and Martin, P. (2009). 'Conflicts and tensions in classroom codeswitching: an introduction'. *International Journal of Bilingual Education and Bilingualism*, 12:2, 117-122.
- Lightbown, P.M. (2008). 'Easy As Pie? Children Learning Languages'. *Concordia Working Papers in Applied Linguistics*, 1, 5–29.
- Littlewood, W., and Yu, B. (2011). 'First language and target language in the foreign language classroom'. *Language Teaching*, 44:1, 64–77.

- Low, G., and Bevertson, S. (2004). 'A systematic review of the impact of ICT on literacy learning in English of learners between 5 and 16, for whom English is a second or additional language'. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Lueder, D.C. (2011). *Involving hard-to-reach partners: creating family/school partnerships*. Lanham: Rowman & Littlefield Education.
- Lugo-Neris, M.J., Jackson, C.W., and Goldstein, H. (2010). 'Facilitating Vocabulary Acquisition of Young English Language Learners'. *Language, Speech, and Hearing Services in Schools*, 41, 314–317.
- Macaro, E. (2001). 'Analysing Student Teachers' Codeswitching in Foreign Language Classrooms: Theories and Decision Making'. *Modern Language Journal*, 85:4, 531-548.
- Macaro, E. (2009) 'Teacher Use of Codeswitching in the Second Language Classroom: Exploring 'Optimal' Use'. In Turnbull, M. and Dailey-O'Cain, J. (eds.), *First Language Use in Second and Foreign Language Learning*. Bristol: Multilingual Matters, pp. 35-49.
- Macaro, E. (2018). *English Medium Instruction*. Oxford: OUP.
- MacCoubrey, S.J., Wade-Woolley, L., Klinger, D., and Kirby, J.R. (2004). 'Early identification of at-risk L2 readers'. *Canadian Modern Language Review*, 61, 11–28.
- MacSwan, J. (2000). 'The threshold hypothesis, semilingualism, and other contributions to a deficit view of linguistic minorities'. *Hispanic Journal of Behavioural Sciences*, 22:1, 3-45.
- Maldonado, J.A. (1994). 'Bilingual special education: Specific learning disabilities in language and reading'. *Journal of Education Issues of Language Minority Students*, 14, 127-147.
- Maldonado, J.R. (1977). 'The effect of the ESEA Title VII program on the cognitive development of Mexican American students'. Unpublished doctoral dissertation, University of Houston, Houston, TX.
- Munafò, M.R., Nosek, B.A., Bishop, D.V.M., Button, K.S., Chambers, C.D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J.J., and Ioannidis, J.P.A. (2017). 'A manifesto for reproducible science'. *Nature Human Behaviour*, 1(1), 0021.
- Mandel, Y. (2018). 'By Degrading Arabic, Israel Has Degraded Arabs'. Haaretz. Friday 20 July. Available at: <https://www.haaretz.com/by-degrading-arabic-israel-has-degraded-arabs-1.6292827> [Accessed 13 Sep 2018].

Marsden, E., Mackey A., and Plonsky, L. (2016). 'The IRIS Repository: Advancing research practice and methodology'. In Mackey, A. and Marsden, E. (eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages*. New York: Routledge, pp. 1-21.

Mathurin, C. and Hoole, C. (1684). *Maturinus Corderius's school-colloquies English and Latine. Divided into several clauses; wherein the propriety of both languages is kept. That children by the help of their mother-tongue, may the better learn to speak Latine in ordinary discourse. There are numbers set down betwixt both, which do shew the place, and natural use of any word or phrase. By Charles Hoole, Mr. of Arts of Lin. Col. Ox. teacher of a private grammar-school, betwixt Goldsmiths-Ally in Red-Cross-street, and Maiden-head Court in Aldersgate-street Lond.* London: the Company of Stationers.

May, T. (2013). 'Immigration Bill' Hansard: House of Commons debates, 22 October c.158. Available at: <https://publications.parliament.uk/pa/cm201314/cmhansrd/cm131022/debtext/131022-0001.htm> [Accessed 18 Dec 2018].

McCambridge, J., Kypri, K. and Elbourne, D. (2014). 'Research participation effects: a skeleton in the methodological cupboard'. *Journal of Clinical Epidemiology*, 67:8, 845–849.

Meisel, J.M., Clahsen, H. and Peinemann, M. (1981). 'On determining developmental stages in natural second language acquisition'. *Studies in Second Language Acquisition*, 3:2, 109-135.

MHCLG (2015). *Indices of deprivation 2015 explorer*. Available at: <http://dclgapps.communities.gov.uk/imd/idmap.html> [Accessed 28 Jun 2018].

Microsoft (2018). *Test your document's readability*. Available at: <https://support.office.com/en-us/article/test-your-document-s-readability-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2> [Accessed 4 Jul 2018].

Milton, J. (2013). 'Measuring the contribution of vocabulary knowledge to proficiency in the four skills'. In Bardel, C., Lindqvist, C., and Laufer, B. (eds.), *L2 Vocabulary Acquisition, Knowledge and Use. New perspectives on assessment and corpus analysis*. Colchester: Eurosla, pp. 57–78.

Moher, D., Hopewell, S., Schulz, K.F., Montori, V., Gøtzscher, P.C., Devereaux, P.J., Elbourne, D., Egger, M. and Altman, D. (2010). 'CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials'. *BMJ*, 340:c896.

- Moher, D., Jadad, A.R., Nichol, G., Penman, M., Tugwell, P., and Walsh, S. (1995). 'Assessing the Quality of Randomized Controlled Trials: An Annotated Bibliography of Scales and Checklists'. *Controlled Clinical Trials*, 16, 62–73.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The PRISMA Group (2009). 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement'. *BMJ*, 339:b2535.
- Monroe, E.E. (1997). 'Effects of mathematical vocabulary instruction on fourth grade students'. *Reading Improvement*, 34, 120–132.
- Moodley, V. (2007). 'Codeswitching in the Multilingual classroom'. *The International Journal of Bilingual Education and Bilingualism*, 100:6, 707-722.
- Moore, D.W. (2008). 'Program development'. In Hinchman, K.A. and Sheridan-Thomas, H.K. (eds.), *Best practices in adolescent literacy instruction*. New York: Guilford, pp. 313–338.
- Moore, P.J. (2013) 'An emergent perspective on the use of the first language in the English-as-a-foreign-language classroom'. *The Modern Language Journal*, 97:1, 239-253.
- Morgan, J.C. (1971). *The effects of bilingual instruction of the English language arts achievement of first grade children*. Doctoral dissertation, Northwestern State University of Louisiana.
- Moyles, J. and Yates, R. (2003). 'What is known about successful models of formative assessment for trainee teachers during school experiences and what constitutes effective practice?' In *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Murphy, L. and Roca de Larios, J. (2010). 'Searching for words: One strategic use of the mother tongue by advanced Spanish EFL writers'. *Journal of Second Language Writing*, 19:2, 61–81.
- Murphy, V.A. (2014). *Learning a second language in the early school years: Trends and contexts*. Oxford: Oxford University Press.
- Murphy, V.A. and Unthiah, A. (2015). *A systematic review of intervention research examining English language and literacy development in children with English as an Additional Language (EAL)*. Available at:  
[https://educationendowmentfoundation.org.uk/public/files/Publications/EAL\\_Systematic\\_review.pdf](https://educationendowmentfoundation.org.uk/public/files/Publications/EAL_Systematic_review.pdf) [Accessed 11 Feb 2019].

Murphy, V.A., Macaro, E., Alba, S., and Cipolla, C. (2014). 'The influence of learning a second language in primary school on developing first language literacy skills'. *Applied Psycholinguistics*, 36:5, 1133–1153.

Nahampun, E. and Sibarani, B., 2014. 'The Effect of Using Frayer Model on Students' Vocabulary Mastery'. *GENRE Journal of Applied Linguistics of FBS Unimed*, 3:1, 1–8.

NALDIC (2013). *EAL pupils 1997-2013*. Available at: [www.naldic.org.uk/Resources/NALDIC/Research and Information/Documents/Copy of EALpupils19972013.xls](http://www.naldic.org.uk/Resources/NALDIC/Research%20and%20Information/Documents/Copy%20of%20EALpupils19972013.xls) [Accessed 6 Feb 2018].

NALDIC (2016). *The percentage of pupils exposed to a language other than English at home has increased. Pupils with English as an additional language 2010-2016, primary and secondary schools*. Available at: [https://drive.google.com/file/d/0B64f5BqGaw\\_8MTIaZlRTQ3JuS1E/view](https://drive.google.com/file/d/0B64f5BqGaw_8MTIaZlRTQ3JuS1E/view) [Accessed 5 Dec 2016].

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, K. and Snowling, M. (1998). 'Semantic processing skills and the development of word recognition: Evidence from children with reading comprehension difficulties'. *Journal of Memory & Language*, 39, 85-101.

Nation, K. and Snowling, M. (2004). 'Beyond phonological skills: Broader language skills contribute to the development of reading'. *Journal of Research in Reading*, 27:4, 342-356.

Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, P. (2003). 'The role of the first language in foreign language learning'. *Asian EFL Journal*, 1, 35–39.

Nocus, I., Florin, A., Guimard, P., and Vernaudeau, J. (2007). 'Effets d'un enseignement en langue kanak sur les compétences oral/écrit en français au cycle 2 en Nouvelle-Calédonie'. *Bulletin de Psychologie*, 491:5, 471–19.

Nova Scotia Department of Education (2006). *Programme d'études Français - Immersion Maternelle à la 3e année Septembre 2006*. Available at: <https://fsl.ednet.ns.ca/sites/default/files/Francais-M-3-complet.pdf> [Accessed 18 Apr 2016].

- Nye, C., Turner, H., and Schwartz, J. (2006). 'Approaches to parent involvement for improving the academic performance of elementary school age children'. *Campbell Systematic Reviews*, 2006:4.
- O'Brien, B.C., Harris, I.B., Beckman, T.J., Reed, D.A., Cook, D.A. (2014). 'Standards for reporting qualitative research: a synthesis of recommendations'. *Academic Medicine*. 89:9, 1245-1251.
- Oakley, A. (1999). 'Paradigm wars: some thoughts on a personal and public trajectory'. *Social Research Methodology*, 2:3, 247-254.
- Oakley, A. (2000). *Experiments in knowing, Gender and method in the social sciences*. New York: The New Press.
- OECD (2012). 'How does class size vary around the world?' *Education Indicators in Focus*, 2012:09 (November), 1-4.
- OECD (2018). *PISA 2015 Results in Focus*. Available at: <https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf> [Accessed 19 Sep 2018].
- Office for National Statistics (2013). *Language in England and Wales: 2011*. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language/articles/languageinenglandandwales/2013-03-04> [Accessed 2 Dec 2016].
- Office for National Statistics (No Date). *Main Language by sex by age*. Available at: <http://www.nomisweb.co.uk/census/2011/lc2104ew> [Accessed 2 Dec 2016].
- Ofsted (2004). *Achievement of Bangladeshi heritage pupils HMI 513*. Available at: [https://dera.ioe.ac.uk/4836/7/Achievement%20of%20Bangladeshi%20heritage%20pupils%20\(PDF%20format\)\\_Redacted.pdf](https://dera.ioe.ac.uk/4836/7/Achievement%20of%20Bangladeshi%20heritage%20pupils%20(PDF%20format)_Redacted.pdf). [Accessed 17 Dec 2018].
- Ofsted (2012a). *School Report St Philip and James' Church of England Aided Primary School Oxford*. Available at: <https://reports.ofsted.gov.uk/provider/files/2069663/urn/123214.pdf> [Accessed 17 Jul 2017].
- Ofsted (2012b). *School Report North Hinksey Church of England Primary School*. Available at: <https://reports.ofsted.gov.uk/provider/files/2320803/urn/123156.pdf> [Accessed 5 Dec 2016].
- Ofsted (2014). *School Report Wolvercote Primary School*. Available at: <https://reports.ofsted.gov.uk/provider/files/2160943/urn/123051.pdf> [Accessed 5 Dec 2016].
- Ofsted (2015). *School Report East Oxford Primary School*. Available at: <https://reports.ofsted.gov.uk/provider/files/2463933/urn/123046.pdf> [Accessed 17 Jul 2017].

Oliver, S. (2015). 'Advantages of concurrent preparation and reporting of systematic reviews of quantitative and qualitative evidence'. *Journal of the Royal Society of Medicine*, 108:3, 108–111.

Otheguy, R., García, O., and Reid, W. (2015). 'Clarifying translanguaging and deconstructing named languages: A perspective from linguistics'. *Applied Linguistics Review*, 6:3, 63–28.

Ovando, C. J. (2003). 'Bilingual Education in the United States: Historical Development and Current Issues'. *Bilingual Research Journal*, 27:1, 1–24.

Oxford Brookes University (2016a). *Why Oxford Brookes?* Available at: <http://www.brookes.ac.uk/international/why-oxford-brookes/about-brookes/> [Accessed 2 Dec 2016].

Oxford Brookes University (2016b). *About Brookes*. Available at: <http://www.brookes.ac.uk/international/why-oxford-brookes/about-brookes/> [Accessed 2 Dec 2016].

Oxford City Council (2013). *First languages spoken by Oxford children (by ward) 2010 School Census*. Available at: [https://www.oxford.gov.uk/downloads/file/1095/first\\_languages\\_spoken\\_by\\_oxford\\_children\\_by\\_ward\\_2010\\_school\\_census](https://www.oxford.gov.uk/downloads/file/1095/first_languages_spoken_by_oxford_children_by_ward_2010_school_census) [Accessed 5 Dec 2016].

Oxford City Council (2016a). *Ethnicity*. Available at: <https://www.oxford.gov.uk/info/20131/population/462/ethnicity> [Accessed 2 Dec 2016].

Oxford City Council (2016b). *International Migration*. Available at: [https://www.oxford.gov.uk/info/20131/population/464/international\\_migration](https://www.oxford.gov.uk/info/20131/population/464/international_migration) [Accessed 2 Dec 2016].

Oxford City Council (2016c). *First or preferred language*. Available at: [https://www.oxford.gov.uk/info/20131/population/463/first\\_or\\_preferred\\_language](https://www.oxford.gov.uk/info/20131/population/463/first_or_preferred_language) [Accessed 5 Dec 2016].

Oxford City Council (2017). *Oxford Profile 2017 Key Facts*. Available at: [https://www.oxford.gov.uk/downloads/file/4036/oxford\\_profile\\_2017](https://www.oxford.gov.uk/downloads/file/4036/oxford_profile_2017) [Accessed 28 Jun 2018].

Oxfordshire County Council (2016). *Schools*. Available at: <https://www.oxfordshire.gov.uk/cms/schools/list/result/Primary/Oxford> [Accessed 5 Dec 2016].



Oxley, E. and de Cat, C. (2018). A systematic review of language and literacy interventions in children and adolescents with English as an additional language (EAL). *OSF Preprints*. October 9. doi:10.31219/osf.io/92s6v.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffman, T., Mulrow, C.D., Shamseer, L., and Moher, D. (2018). *Updating the PRISMA reporting guideline for systematic reviews and meta-analyses: study protocol*. Available at: <https://osf.io/2v7mk/> [Accessed 27 Aug 2018].

Palmer, J., Boon, R.T., and Spencer, V.G. (2014) 'Effects of Concept Mapping Instruction on the Vocabulary Acquisition Skills of Seventh-Graders With Mild Disabilities: A Replication Study'. *Reading and Writing Quarterly*, 30:2, 165–182.

Paradowski, M.B. (2007). 'Foreign-language grammar instruction via the mother tongue'. In Boers, F., Darquennes, J., and Temmerman, R. (eds.), *Multilingualism and Applied Comparative Linguistics, Vol. Comparative Considerations in Second and Foreign Language Instruction*. Newcastle-upon-Tyne: Cambridge Scholars Press, pp. 151–167.

Paribakht, T.S., and Wesche, M. (1997). 'Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition'. In Coady, J. and Huckin, T. (eds.), *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge: Cambridge University Press, pp. 174–200.

Parke, T., Drury, R., Kenner, C., and Robertson, L.H. (2002). 'Revealing invisible worlds: Connecting the mainstream with bilingual children's home and community learning'. *Journal of Early Childhood Literacy*, 2:2, 195–220.

Pattison High School (2017). *English Only Policy*. Available at: <http://www.pattisonhighschool.ca/english-only-policy/> [Accessed 8 Aug 2018].

Paulston, C.B. (1975). 'Ethnic relations and bilingual education: accounting for contradictory data'. *Working Papers on Bilingualism*, No. 6., Toronto: OISE.

Perozzi, J.A. and Chavez Sanchez, M.L. (2003). 'The Effect of Instruction in L1 on Receptive Acquisition of L2 for Bilingual Children with Language Delay'. *Language, Speech, and Hearing Services in Schools*, 23, 348–352.

Peters, C. (1974). 'A Comparison between the Frayer Model of Concept Attainment and the Textbook Approach to Concept Attainment (Abstract)'. *Reading Research Quarterly*, 10:2, 252–254.

- Petrosino, A., Boruch, R., Rounding, C., McDonald, S. and Chalmers, I. (2000). 'The Campbell Collaboration Social, Psychological, Educational and Criminological Trials Register (C2-SPECTR) to Facilitate the Preparation and Maintenance of Systematic Reviews of Social and Educational Interventions'. *Evaluation and Research in Education*, 14:3-4, 206–219.
- Petrovic, J.E. and Olmstead, S. (2001). 'Language, power, and pedagogy: bilingual children in the crossfire, by J. Cummins'. *Bilingual Research Journal*, 25:3, 405-412.
- Petticrew, M. and Roberts, H. (2006). *Systematic reviews in the social sciences A practical guide*. Oxford: Blackwell.
- Piaget, J. (1923). "*Langage et pensée chez l'enfant*" (3e éd. 1948 revue et avec un nouvel avant-propos et un nouveau chapitre II inséré utgave bind). Neuchâtel: Delachaux et Niestlé.
- Plante, A. J. (1976). *A study of effectiveness of the Connecticut "Pairing" model of bilingual/bicultural education*. Hamden, CT: Connecticut Staff Development Cooperative.
- Prideaux, D. (2002). 'Researching the outcomes of educational interventions: a matter of design'. *BMJ*, 324, 126–127.
- QCA (2000). *Curriculum guidance for the foundation stage*. London: Qualifications and Curriculum Authority.
- Ramirez, J., Pasta, D.J., Yuen, S., Billings, D.K., and Ramey, D.R. (1991). *Final report: Longitudinal study of structural immersion strategy, early-exit, and late-exit transitional bilingual education programs for language-minority children (Report to the U.S. Department of Education)*. San Mateo, CA: Aguirre International.
- Random.org (1998-2017). *Random.org True random number service*. Available at: [www.random.org](http://www.random.org)
- Reljić, G., Ferring, D., and Martin, R. (2015). 'A meta-analysis on the effectiveness of bilingual programs in Europe'. *Review of Educational Research*, 85:1, 92–128.
- ResearchED (2018a). *Our Story*. Available at: <https://researched.org.uk/about/our-story/> [Accessed 30 Jul 2018].
- ResearchED (2018b). *Our Mission*. Available at: <https://researched.org.uk/about/our-story/> [Accessed 30 Jul 2018].
- Rochford, D. (2015a). *Statement on the interim recommendations of the Rochford Review December 2015*. London: DfE.

- Rochford, D. (2015b). *Statement on the interim recommendations of the Rochford Review (Update: February 2016)*. Available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/498204/Interim\\_recommendations\\_of\\_the\\_Rochford\\_Review.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/498204/Interim_recommendations_of_the_Rochford_Review.pdf). [Accessed 6 Aug 2018].
- Rolstad, K., Mahoney, K., and Glass, G.V. (2005). 'The Big Picture: A Meta-Analysis of Program Effectiveness Research on English Language Learners'. *Educational Policy*, 19:4, 572–594.
- Rott, S., Williams, J., and Cameron, R. (2002). 'The effect of multiple-choice L1 glosses and input-output cycles on lexical acquisition and retention'. *Language Teaching Research*, 6:3, 183–222.
- Roussel, S., Joulia, D., Tricot, A., and Sweller, J. (2017). 'Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach'. *Learning and Instruction*, 52, 69-78.
- Ryan, A.M. (2005). 'The effectiveness of the Manchester Even Start Program in improving literacy outcomes for preschool Latino students'. *Journal of Research in Childhood Education*. 20:1, 15–26.
- Ryan, C. (2013). *Language use in the United States: 2011. American community survey reports 2013, ACS-22*. Available at: <https://www.census.gov/prod/2013pubs/acs-22.pdf> [Accessed 22 Sep 2016].
- Saldate, M., Mishra, S.P., and Medina, M. (1985). 'Bilingual instruction and academic achievement: A longitudinal study'. *Journal of Instructional Psychology*, 12:1, 24-30.
- Sánchez, L.Z. (2004). *Effects of Parent Participation Using First Language Curriculum Materials on the English Reading Achievement and Second-Language Acquisition of Hispanic Students*. Unpublished PhD Thesis. Lehigh University, Pennsylvania, USA.
- Schöpfel, J. (2010). 'Towards a Prague Definition of Grey Literature. Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues'. Prague, 6-7 December 2010, Czech Republic. pp. 11-26.
- Schulte, P.A. (1991). 'Ethical issues in the communication of results'. *Journal of Clinical Epidemiology*, 44:Supplement 1, 57-61.
- Schulz, K.F., Altman, D.G., Moher, D., for the CONSORT Group (2010). 'CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials'. *Lancet*. 375:9721, 1136.
- Scimago Lab (2007-2018). *Scimago Journal & Country Rank*. Available at: <https://www.scimagojr.com/journalrank.php?category=3304> [Accessed 13 Sep 2018].

Scott, S. (2016). 'Schools must report EAL data differently from September - are you ready?' *Schools Week*, Saturday 9 July. Available at: <http://schoolsweek.co.uk/eal-is-this-a-change-for-the-better/> [Accessed 17 Jul 2017].

Scott, V.M. and de la Fuente, M.J. (2008). 'What's the problem? Use of the LI consciousness-raising, L2 learners'. *The Modern Language Journal*, 92:1, 100–113.

Senn, S. (1993). *Cross-over trials in clinical research*. Chichester: John Wiley and Sons.

Shadish, W., Cook, T., and Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton, Mifflin and Company.

Sieh, Y.-C. (2008). 'A possible role for the first language in young learners' processing and storage of foreign language vocabulary'. *Annual Review of Education, Communication, and Language Sciences*, 5, 136–160.

Simons, G.F. and Fennig, C.D. (eds.) (2017). *Ethnologue: Languages of the World, Twentieth edition*. Dallas, TX: SIL International.

Slavin, R.E. (1986). 'Best-evidence synthesis: an alternative to meta-analytic and traditional reviews'. *Educational Researcher*, 15:9, 5–11.

Slavin, R.E. and Cheung, A. (2005). 'A synthesis of research on language of reading instruction for English language learners'. *Review of Educational Research*, 75:2, 247–284.

Slavin, R.E., Lake, C., Chambers, B., Cheung, A., and Davis, S. (2009). 'Effective Reading Programs for the Elementary Grades: A Best-Evidence Synthesis'. *Review of Educational Research*, 79:4, 1391–1466.

Smith, S.A. and Murphy, V.A. (2015). 'Measuring productive elements of multi-word phrase vocabulary knowledge among children with English as an additional or only language'. *Reading and Writing*. 28:3, 347-369.

Smithells, R.W. (1975). 'Iatrogenic hazards and their effects'. *Postgraduate Medical Journal*, 15, 39-52.

Sneddon, R. (2000). 'Language and Literacy: Children's Experiences in Multilingual Environments'. *International Journal of Bilingual Education and Bilingualism*, 3:4, 265–282.

Souvignier, E., Duzy, D., Glück, D., Pröscholdt, M.V., and Schneider, W. (2012). 'Fostering preschoolers phonological awareness in children with German as a second language: Effects of

programs in the native language and in the second language'. *Zeitschrift fur Entwicklungspsychologie und Padagogische Psychologie*, 44:1, 40-51.

Sparks, R., and Ganschow, L. (1991). 'Foreign Language Learning Differences: Affective or Native Language Aptitude Differences?' *Modern Language Journal*, 75:1, 3-16.

Sparks, R., Patton, J., Ganschow, L., Humbach, N., and Javorsky, J. (2008). 'Early first-language reading and spelling skills predict later second-language reading and spelling skills'. *Journal of Educational Psychology*, 100, 162-174.

Stanat, P., and Christensen, G. (2006). *Where Immigrant Students Succeed*. Available at: [https://www.phil-fak.uni-duesseldorf.de/ew/bf/bf\\_veranstaltungen/.../9806021E.pdf](https://www.phil-fak.uni-duesseldorf.de/ew/bf/bf_veranstaltungen/.../9806021E.pdf) [Accessed 26 Oct 2017].

Steele, J.L., Slater, R.O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., and Bacon, M. (2017a). 'Effects of Dual-Language Immersion Programs on Student Achievement'. *American Educational Research Journal*, 54:Supplement 1, 282S–306S.

Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., and Bacon, M. (2017b). 'Dual-Language Immersion Programs Raise Student Achievement in English'. Santa Monica, CA: RAND Corporation. Available at: [https://www.rand.org/pubs/research\\_briefs/RB9903.html](https://www.rand.org/pubs/research_briefs/RB9903.html). [Accessed 11 Apr 2018].

Stewart, H. and Mason, R. (2016). 'Nigel Farage's anti-migrant poster reported to police'. *The Guardian* (Thursday June 16). Available at: <https://www.theguardian.com/politics/2016/jun/16/nigel-farage-defends-ukip-breaking-point-poster-queue-of-migrants> [Accessed 17 Dec 2018].

Storch, N. and Wigglesworth, G. (2003). 'Is there a role for the use of the L1 in an L2 setting?' *TESOL Quarterly*, 37:4, 760-770.

Storey (2016). 'Education (Pupil Information) (England) (Miscellaneous Amendments) Regulations 2016', Hansard: House of Lords Motion to Regret, 31 October, 776, c.503. Available at: [https://hansard.parliament.uk/Lords/2016-10-31/debates/6D06F8D5-7709-43DF-87ED-33CBBC7324FF/Education\(PupilInformation\)\(England\)\(MiscellaneousAmendments\)Regulations2016](https://hansard.parliament.uk/Lords/2016-10-31/debates/6D06F8D5-7709-43DF-87ED-33CBBC7324FF/Education(PupilInformation)(England)(MiscellaneousAmendments)Regulations2016). [Accessed 20 July 2018].

Strand, S., Malmberg, L. and Hall, J. (2015). *English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database*. Available at:

<https://educationendowmentfoundation.org.uk/evidence-summaries/evidence-reviews/english-as-an-additional-language-eal/> [Accessed 11 Feb 2019].

Strand, S. and Hessel, A. (2018). *English as an Additional Language, proficiency in English and pupils' educational achievement: An analysis of Local Authority data*. Available at: <https://www.bell-foundation.org.uk/wp-content/uploads/2018/10/EAL-PIE-and-Educational-Achievement-Report-2018-FV.pdf>. [Accessed 5 Dec 2018].

Strohmeier, B., and McGrail, L. (1988). *On FOCUS: Photographs and Writings by Students*. Boston, MA: El Centro Del Cardenal.

Studies in Second Language Acquisition (2017). *Instructions for contributors*. Available at: <https://www.cambridge.org/core/journals/studies-in-second-language-acquisition/information/instructions-contributors> [Accessed 14 Sep 2018].

Sunderman, G., and Kroll, J. F. (2006). 'First language activation during second language lexical processing: an investigation of lexical form, meaning, and grammatical class'. *Studies in Second Language Acquisition*, 28:3, 387–422.

Swain, M. (1995). 'Three functions of output in second language learning'. In Cook, G. and Seidlhofer, B. (eds.), *Principle and practice in applied linguistics: studies in honour of H.G. Widdowson*. Oxford: Oxford University Press, pp. 125–44.

Swain, M. (1999). 'Language production in SLA: from social to cognitive and back again'. Plenary address at SLRF, University of Minnesota.

Swain, M. (2000). 'The output hypothesis and beyond: mediating acquisition through collaborative dialogue'. In Lantolf, J.P. (ed.), *Sociocultural theory and second language learning*. Oxford: Oxford University Press, pp. 97–114.

Swain, M., and Lapkin, S. (1998). 'Interaction and second language learning: two adolescent French immersion students working together'. *Modern Language Journal*, 82:iii, 320–37.

Swain, M. and Lapkin, S. (2000a). 'Task-based second language learning: the uses of the first language'. *Language Teaching Research*, 4:3, 251–274.

Swain, M. and Lapkin, S. (2000b). 'Focus on form through collaborative dialogue: exploring task effects'. In Bygate, M., Skehan, P., and Swain, M. (eds.), *Researching pedagogic tasks: second language learning, teaching and testing*. Harlow, Essex: Longman, pp. 99–118.

- Swan, M. (1985a). 'A critical look at the communicative approach (2)'. *ELT Journal*, 39:1, 2-12.
- Swann, M. (1985b). *Education for All. The report of the Committee of Inquiry into the Education of Children from Ethnic Minority Groups*. London: HMSO.
- The Bell Foundation (2019). *The Bell Foundation*. Available at: <https://www.bell-foundation.org.uk/> [Accessed 1 Jan 2019].
- The Sutton Trust (2012-1018). *Sutton Trust – improving social mobility*. Available at: <https://www.suttontrust.com/> [Accessed 1 Jan 2019].
- Thomas, B.H., Ciliska, D., Dobbins, M. and Micucci, S. (2004). 'A Process for Systematically Reviewing the Literature: Providing the Research Evidence for Public Health Nursing Interventions'. *World Views on Evidence-Based Nursing*, 1:3, 176–184.
- Thomas, G. (2013a). 'No one can control for a sense of when 4-3-3 might turn the game'. *Times Higher Education*, Sunday 17 February. Available at: <https://www.timeshighereducation.com/comment/opinion/no-one-can-control-for-a-sense-of-when-4-3-3-might-turn-the-game/2001371.article> [Accessed 30 Jul 2018].
- Thomas, G. (2013b). *How to Do Your Research Project. A Guide for Students in Education and Applied Social Sciences (Second Edition)*. London: SAGE.
- Thomas, J., Harden, A., Oakley, A., Oliver, S., Sutcliffe, K., Rees, R., Brunton, G., Kavanagh, J. (2004). 'Integrating qualitative research with trials in systematic reviews'. *BMJ*, 328, 1010-1012.
- Thompson, K.D. (2015). 'English Learners' Time to Reclassification: An Analysis'. *Educational Policy*, 31:3, 330–363.
- Tikunoff, W.J. (1985). *Applying Significant Bilingual Instructional Features in the Classroom*. Rosslyn, VA: National Clearinghouse for Bilingual Education.
- Tingley, P.A., Dore, K.A., Lopez, A., Parsons, H., Campbell, E., Bird, E.K., and Cleave, P. (2004). 'A comparison of phonological awareness skills in early French immersion and English children'. *Journal of Psycholinguistic Research*, 33, 263–287.
- Tonzar, C., Lotto, L. and Job, R. (2009). 'L2 Vocabulary Acquisition in Children: Effects of Learning Method and Cognate Status'. *Language Learning*, 59:3, 623–646.
- Torgerson, C. (2003) *Systematic Reviews*. London: Continuum.

Torgerson, C.J., Wiggins, A., Torgerson, D.J., Ainsworth, H., Barmby, P., Hewitt, C., Jones, K., Hendry, V., Askew, M., Bland, M., Coe, R., Higgins, S., Hodgen, J., Hulme, C. and Tymms, P. (2011). *Research Report DFE-RR091a. Every Child Counts: the independent evaluation Technical Report*. London: Department for Education.

Torgerson, D. and Torgerson, C. (2008). *Designing Randomised Trials in Health, Education and the Social Sciences: An Introduction*. Basingstoke: Palgrave MacMillan.

Travis, A. (2013). 'Immigration bill: Theresa May defends plans to create 'hostile environment''. *The Guardian*, Thursday 10 October. Available at:

<https://www.theguardian.com/politics/2013/oct/10/immigration-bill-theresa-may-hostile-environment> [Accessed 18 Dec 2018].

Turnbull, M. and Daily-O'Cain, J. (2009). *First Language Use in Second and Foreign Language Learning*. Bristol: Multilingual Matters.

Unbound Philanthropy (no date). *Unbound*. Available at: <https://www.unboundphilanthropy.org/> [Accessed 1 Jan 2019].

United Nations (1990). *The United Nations Convention on the Rights of the Child*. Available at: [https://downloads.unicef.org.uk/wp-content/uploads/2010/05/UNCRC\\_united\\_nations\\_convention\\_on\\_the\\_rights\\_of\\_the\\_child.pdf](https://downloads.unicef.org.uk/wp-content/uploads/2010/05/UNCRC_united_nations_convention_on_the_rights_of_the_child.pdf) [Accessed 13 Sep 2018].

United Nations (2017). *International Mother Language Day 21 February*. Available at: <http://www.un.org/en/events/motherlanguageday/> [Accessed 13 Sep 2018].

University of Oxford (2016a). *Oxford's International Profile*. Available at: <https://www.ox.ac.uk/about/international-oxford/oxfords-international-profile?wssl=1> [Accessed 2 Dec 2016].

University of Oxford (2016b). *International Students*. Available at: <https://www.ox.ac.uk/students/new/international?wssl=1> [Accessed 2 Dec 2016].

Valdés, G. and Figueroa, R. (1994). *Bilingualism and testing A special case of bias*. Norwood, NJ: Ablex Publishing Corporation.

van Gelderen, A., Schoonen, R., Stoel, R.D., de Glopper, K., and Hulstijn, J. (2007). 'Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components'. *Journal of Educational Psychology*, 99:3, 477–491.



van Weijen, D., van den Bergh, H., Rijlaarsdam, G., and Sanders, T. (2009). 'L1 use during L2 writing: An empirical study of a complex phenomenon'. *Journal of Second Language Writing*, 18:4, 235–250.

von Elm, E., Altman, D.G., Egger, M., Pocock, S.J., Gøtzsche, P.C., Vandenbroucke, J.P. (2007). 'The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies'. *Lancet*. 370:9596, 1453-1457.

Wallace, C., Mallows, D., Abley, J., Conteh, J., Mehmedbegovic, D. and West, R. (2009). *English as an Additional Language (EAL) provision in schools – 10 case studies*. London: IOE.

Walters, J.E. (1931). 'Seniors as Counselors'. *The Journal of Higher Education*, 2:8, 446–448.

Walters, K. and Gunderson, L. (1985). 'Effects of parent volunteers reading first language (L1) books to ESL students'. *The Reading Teacher*, 39:1, 66–69.

Wang, W. and Wen, Q. (2002). 'L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers'. *Journal of Second Language Writing*, 11:3, 225-246.

Wang, W. and Wen, Q. (2002). 'L1 use in the L2 composing process: an exploratory study of 16 Chinese EFL writers'. *Journal of Second Language Writing*, 11:3, 225-246.

Watson, F. (1908). *The English grammar schools to 1660: their curriculum and practice*. Cambridge: Cambridge University Press. Available at: <https://archive.org/details/cu31924082458229> [Accessed 19 Dec 2018].

Weinreich, M. (1945). 'Der YIVO un di Problemen fun Undzer Tsayt'. *YIVO-Bleter*, 25:1, 13.

What Works Clearinghouse (2014). *Procedures and Standards Handbook, Version 3.0*. Available at: <http://ies.ed.gov/ncee/wwc/DocumentSum.aspx?sid=19> [Accessed 31 May 2016].

What Works Clearinghouse (2015a) *WWC Intervention Report A summary of findings from a systematic review of the evidence Everyday Mathematics®*. Available at: [http://ies.ed.gov/ncee/wwc/pdf/intervention\\_reports/wwc\\_everydaymath\\_111015.pdf](http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_everydaymath_111015.pdf) [Accessed 31 May 2016].

What Works Clearinghouse (2015b). *WC Intervention Report A summary of findings from a systematic review of the evidence Credit Recovery Programs*. Available at: [http://ies.ed.gov/ncee/wwc/pdf/intervention\\_reports/wwc\\_credit\\_050515.pdf](http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_credit_050515.pdf) [Accessed 31 May 2016].

Whittaker, F. and Camden, B. (2016). 'Pupils who were not white British told to send in birthplace data'. *Schools Week*, Friday 23 September. Available at: <https://schoolsweek.co.uk/pupils-who-were-not-white-british-told-to-send-in-birthplace-data/> [Accessed 17 Dec 2018].

Wijekumar, K., Meyer, B.J.F. and Lei, P. (2014). 'Improving content area reading comprehension with 4-6th grade Spanish ELLs using web-based structure strategy instruction'. In *11th International Conference on Cognition and Exploratory Learning in Digital Age*. 158-164.

Wiliam, D. (2018a). *Nothing. It's just that there is very little evidence that it works. Doing things that might work, while not doing things that are known to work is, in my view, unprofessional, like withholding effective treatments for illnesses in order to research others....* Available at: <https://twitter.com/dylanwiliam/status/1028967287372632069> [Accessed 16 Aug 2018].

Wiliam, D. (2018b). *Easy. Compare new approaches to the best (i.e., most cost-effective) available approaches, rather than to just "business as usual."* Available at: <https://twitter.com/dylanwiliam/status/1029379287000989696> [Accessed 16 Aug 2018].

Wixson, K.K. (1986). 'Vocabulary Instruction and Children's Comprehension of Basal Stories'. *Reading Research Quarterly*, 21:3, 317-329.

World Medical Association (1975). *Declaration of Helsinki*. Available at: <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/> [Accessed 16 Aug 2018].

Yeh, R.W., Valsdottir, L.R., Yeh, M.W., Shen, C., Kramer, D.B., Strom, J.B., Secemsky, E.A., Healy, J.L., Domeier, R.M., Kazi, D.S., and Nallamotheu, B.K. (2018). 'Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial'. *BMJ*, 363:k5094.

Yiakoumetti, A. (2006). 'A Bidialectal Programme for the Learning of Standard Modern Greek in Cyprus'. *Applied Linguistics*, 27:2, 295–317.

Yiakoumetti, A. (2007). 'Choice of classroom language in bidialectal communities: to include or to exclude the dialect?' *Cambridge Journal of Education*, 37:1, 51–66.

Yiakoumetti, A. (2011). 'Re-Invigorating Teacher-Training Programmes in Linguistic Variation'. In Saleh, I.M. and Khine, M.S. (eds.), *Teaching Teachers*. New York: Nova Science Publishers, Inc. pp. 195–214

Zhang, D. and 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 effect'. *Reading and Writing*, 25:5, 1195-1216.

Zhang, D. and Koda, K. (2013). 'Morphological awareness and reading comprehension in a foreign language: A study of young Chinese EFL learners'. *System*, 41:4, 901-913.

Zhang, D. and Koda, K. (2014). 'Awareness of derivation and compounding in Chinese-English biliteracy acquisition'. *International Journal of Bilingual Education and Bilingualism*, 17:1, 55-73.

# APPENDICES

## **Systematic review tools**

**Appendix A:** Coding sheet for studies identified as eligible for the systematic review

**Appendix B:** Adapted EPPHP Risk of Bias Tool

## **Informed consent**

**Appendix C:** Participant information sheet for parents

**Appendix D:** Participant information sheet for children

**Appendix E:** Parental consent form

## **Intervention materials**

**Appendix F:** Language background questionnaire

**Appendix G:** Expressive vocabulary test

**Appendix H:** Receptive vocabulary test

**Appendix I:** Participant exercise book

**Appendix J:** Target items of vocabulary: Long list, Reduced list, Final items

# CODING SHEET

Reference:

Item	Response	Guidance
<b>Administrative</b>		
1. How was the report identified?		<i>Was the report identified through an electronic database search, iterative forward/backward citation search, already known to reviewer etc.</i>
2. What was the language of publication?		<i>State the language in which the full text of the study is written.</i>
<b>Contextual information about the study</b>		
3. Where was the study conducted?		<i>Collect information about the geographical location of the study in as much detail as is provided by the authors. If the author states the name of the city and country in which the study was carried out, note that. If the authors only indicate the characteristics of the location (for example "An urban area of Northern Spain") note them here. If this description is based on reviewer interpretation rather than information directly from the</i>

		<i>paper, say so. If location is not stated and cannot be inferred from other information, code as 'not stated'.</i>
4. In what type of school was the study conducted?		<i>Did the study take place in a mainstream school, language institute, a complimentary/community school, a pre-school programme, etc.? If the authors do not name the type of school used, reviewer should try to infer the programme type from contextual information, and indicate that this is reviewer interpretation. If school type is not stated and cannot be inferred from other information, code as 'not stated'.</i>
5. What term best characterizes the language teaching/programme within the institute?		<i>If the authors specifically state the type of language programme, use their term(s). These may include, for example: No special provision, ESL/ESOL programme, immersion, or MFL. If the authors do not name the type of language programme used, reviewer should try to infer the programme type from contextual information, and indicate that this is reviewer interpretation. If programme type is not stated and cannot be inferred from other information, code as 'not stated'.</i>
6. Who were the participants and how are they characterised?		<i>Note information that is reported about the sex, age, socio economic status, linguistic backgrounds, etc. of the participants.</i>
<b><i>Descriptive information about the intervention(s)</i></b>		
7. Research question		<i>What was the research question of the study (or part of the study that pertains to this review)? If this is not articulated by the authors and has been interpreted by the reviewer, say so.</i>
8. What was the L1-mediated intervention.		<i>Give a basic description of the intervention that made use of the participants L1.</i>

9. What was the L1-mediated intervention compared to.		<i>If the L1-mediated intervention was compared with an alternative intervention, describe the alternative intervention. If there was no comparator, say so.</i>
10. Who delivered the intervention and to how many participants at a time?		<i>Note who delivered the intervention (e.g. usual teacher, teaching assistant, researcher, parent, or combinations of these) and whether this was to one child at a time, small groups of children, whole classes etc.</i>
11. What language domains and/or skills constituted the principal independent and dependent variables?		<i>The intervention will make use of L1 from one or more of the four language domains: reading, writing, speaking and listening. The outcome will be measured in terms of the same domains or of different domains. In addition, the output may specify a specific skill or aspect of that larger domain, such as spelling, comprehension, or vocabulary use. By way of illustration, a study that assesses the effects of discussing a topic with students in either their L1 or L2 on the lexical density of their subsequent L2 written account of the content of that discussion would be coded as “Input: Speaking, Output: Writing; lexical density”. Another study, assessing the effects of drafting an essay either in L1 or L2 on the linguistic complexity of a subsequent L2 write-up would be coded as “Input: Writing, Output: Writing; linguistic complexity”.</i>
12. What were the languages used in the study?		<i>State the main language of the classroom (the L2), and the other language(s) used in the intervention (the L1). For example, an intervention in a school in the UK that incorporated students’ use of Polish would be coded as L2=English and L1=Polish.</i>
<b>Methodology and Design</b>		
13. Study Design		<i>What study design was used? Refer to the EPHPP Quality Assessment Tool for Quantitative Studies Dictionary for names of possible designs and descriptions of them.</i>

14. Did the outcome measures reflect linguistic proficiency or academic attainment?		<i>Were the effects of the intervention measured in terms of the pupils' use or understanding of L2 or with how well the pupils demonstrated understanding of curriculum content, or both? If the outcome measure(s) was something else, specify.</i>
15. Number of Participants		<i>How many participants took part in the study?</i>
16. Duration		<i>Over what length of time did the study take place?</i>
17. Quality Assessment Rating		<i>Use the EPHPP Tool to assess the quality of the study and note the overall strength of the study here (strong, medium, or weak)</i>



## Appendix B: Adapted EPPHP Risk of Bias Tool

### QUALITY ASSESSMENT TOOL FOR QUANTITATIVE STUDIES

Reference:

#### COMPONENT RATINGS

##### A) SELECTION BIAS

(Q1) Are the individuals selected to participate in the study likely to be representative of the target population?

- 1 Very likely
- 2 Somewhat likely
- 3 Not likely
- 4 Can't tell

(Q2) What percentage of selected individuals agreed to participate?

- 1 80 - 100% agreement
- 2 60 – 79% agreement
- 3 less than 60% agreement
- 4 Not applicable
- 5 Can't tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

##### B) STUDY DESIGN

Indicate the study design

1. Randomized controlled trial
2. Non-randomized comparison (comparison groups prospectively generated by a means other than random allocation)
3. Cohort analytic (two group pre + post)
4. Case-control
5. Cohort (one group pre + post (before and after))
6. Interrupted time series
7. Other specify \_\_\_\_\_
8. Can't tell

Was the study described as randomized? If NO, go to Component C.

No Yes

If Yes, was the method of randomization described? (See dictionary)

No Yes

If Yes, was the method appropriate? (See dictionary)

No Yes

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

**C) CONFOUNDERS**

**(Q1) Were there known important differences between groups prior to the intervention?**

- 1 Yes
- 2 No
- 3 Can't tell

**The following are examples of confounders:**

- 1 Race
- 2 Sex
- 3 Marital status/family
- 4 Age
- 5 SES (income or class)
- 6 Education
- 7 Health status
- 8 Pre-intervention score on outcome measure

**(Q2) If yes, indicate the percentage of relevant confounders that were controlled (either in the design (e.g. stratification, matching) or analysis)?**

- 1 80 – 100% (most)
- 2 60 – 79% (some)
- 3 Less than 60% (few or none)
- 4 Can't Tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

**D) BLINDING**

**(Q1) Was (were) the outcome assessor(s) aware of the intervention or exposure status of participants?**

- 1 Yes
- 2 No
- 3 Can't tell

**(Q2) Were the study participants aware of the research question?**

- 1 Yes
- 2 No
- 3 Can't tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

**E) DATA COLLECTION METHODS**

**(Q1) Were data collection tools shown to be valid?**

- 1 Yes
- 2 No
- 3 Can't tell

**(Q2) Were data collection tools shown to be reliable?**

- 1 Yes
- 2 No
- 3 Can't tell

RATE THIS SECTION	STRONG	MODERATE	WEAK
See dictionary	1	2	3

**F) WITHDRAWALS AND DROP-OUTS**

**(Q1) Were withdrawals and drop-outs reported in terms of numbers and/or reasons per group?**

- 1 Yes
- 2 No
- 3 Can't tell
- 4 Not Applicable (i.e. one time surveys or interviews)

**(Q2) Indicate the percentage of participants completing the study. (If the percentage differs by groups, record the lowest).**

- 1 80% - 100%
- 2 60% - 79%
- 3 Less than 60%
- 4 Can't tell
- 5 Not Applicable (i.e. Retrospective case-control)

RATE THIS SECTION	STRONG	MODERATE	WEAK	
See dictionary	1	2	3	Not Applicable

**G) INTERVENTION INTEGRITY**

**(Q1) What percentage of participants received the allocated intervention or exposure of interest?**

- 1 80% - 100%
- 2 60% - 79%
- 3 Less than 60%
- 4 Can't tell

**(Q2) Was the consistency of the intervention measured?**

- 1 Yes
- 2 No
- 3 Can't tell

**(Q3) Is it likely that subjects received an unintended intervention (contamination or co-intervention) that may influence the results?**

- 4 Yes
- 5 No
- 6 Can't tell

**H) ANALYSES**

**(Q1) Indicate the unit of allocation (circle one)**

Local authority :: school :: class/teaching group :: pupil

**(Q2) Indicate the unit of analysis (circle one)**

Local authority :: school :: class/teaching group :: pupil

**(Q3) Are the statistical methods appropriate for the study design?**

- 1 Yes
- 2 No
- 3 Can't tell

**(Q4) Is the analysis performed by intervention allocation status (i.e. intention to treat) rather than the actual intervention received?**

- 1 Yes
- 2 No
- 3 Can't tell

**GLOBAL RATING**

**COMPONENT RATINGS**

Please transcribe the information from the gray boxes on pages 1-4 onto this page. See dictionary on how to rate this section.

<b>A</b>	<b>SELECTION BIAS</b>	<b>STRONG</b>	<b>MODERATE</b>	<b>WEAK</b>
		<b>1</b>	<b>2</b>	<b>3</b>
<b>B</b>	<b>STUDY DESIGN</b>	<b>STRONG</b>	<b>MODERATE</b>	<b>WEAK</b>
		<b>1</b>	<b>2</b>	<b>3</b>
<b>C</b>	<b>CONFOUNDERS</b>	<b>STRONG</b>	<b>MODERATE</b>	<b>WEAK</b>
		<b>1</b>	<b>2</b>	<b>3</b>
<b>D</b>	<b>BLINDING</b>	<b>STRONG</b>	<b>MODERATE</b>	<b>WEAK</b>
		<b>1</b>	<b>2</b>	<b>3</b>
<b>E</b>	<b>DATA COLLECTION METHOD</b>	<b>STRONG</b>	<b>MODERATE</b>	<b>WEAK</b>
		<b>1</b>	<b>2</b>	<b>3</b>
<b>F</b>	<b>WITHDRAWALS AND DROPOUT</b>	<b>STRONG</b>	<b>MODERATE</b>	<b>WEAK</b>
		<b>1</b>	<b>2</b>	<b>3</b>
				<b>NOT APPLICABLE</b>

**GLOBAL RATING FOR THIS PAPER (circle one):**

<b>1</b>	<b>STRONG</b>	<b>(no WEAK ratings)</b>
<b>2</b>	<b>MODERATE</b>	<b>(one WEAK rating)</b>
<b>3</b>	<b>WEAK</b>	<b>(two or more WEAK ratings)</b>

**With both reviewers discussing the ratings:**

**Is there a discrepancy between the two reviewers with respect to the component (A-F) ratings?**

No Yes

**If yes, indicate the reason for the discrepancy**

- 1 Oversight
- 2 Differences in interpretation of criteria

3 Differences in interpretation of study

**Final decision of both reviewers (circle one):**

- 1 STRONG
- 2 MODERATE
- 3 WEAK

## Appendix C: Participant information sheet for parents

### WHAT ARE THE EFFECTS OF ENCOURAGING MULTILINGUAL CHILDREN TO USE THEIR HOME LANGUAGES TO LEARN NEW ENGLISH VOCABULARY?

#### Introduction

This is an invitation for your child to take part in a doctoral research project being conducted at Oxford Brookes University. The project aims to understand how teachers can help children who know more than one language to do well at school.

In some schools, teachers are encouraged to give multilingual children opportunities to use home languages (sometimes called mother-tongues) when they are learning. In other schools, teachers are told that multilingual children should use only English for their school work. There are plausible reasons why both pieces of advice might be correct. Unfortunately, there has not been very much research that would help teachers in England to understand which advice is most likely to help their pupils.

To help teachers decide what approach they will take with multilingual children, this research project will try to find out what happens when multilingual children are given opportunities to use their home languages when learning the meaning of English words.

We are inviting all children who know more than one language in Years 4, 5 and 6 at your child's school to volunteer to help us.

#### Who is conducting this research?

This project is part of a PhD project at Oxford Brookes University. Hamish Chalmers is what is known as the Principal Investigator on the project. That means he will be the person who will work directly with the schools and children involved in it. Hamish is a very experienced primary school teacher who has worked extensively with bilingual children.

Hamish will be supervised by Dr Androula Yiakoumetti, an experienced educational researcher at Oxford Brookes University.

The project has been approved by Oxford Brookes University's Research Ethics Committee.

Contact details for everyone mentioned here are at the end of this leaflet. You are welcome to contact them if you would like more information.

#### Why are we doing this research?

Multilingual children make up 20% of the children in schools in England. This is a large and important group of children. If we understand how multilingual children use their other languages, we can help teachers to provide learning activities that help those children to do well at school. Unfortunately, there is very little research in England on using home languages in schools. We are doing this research so to add to the small amount of research that exists so that we can understand how to teach multilingual children in England better.

#### What will your child do to help us answer our questions?

We will compare two ways of teaching English vocabulary. One way will use English for the whole lesson. The other way will use the children's other languages as well as English for parts of the lesson. We will use an unbiased, or 'fair', method to decide which children are in which group.

Children will be taught 'academic vocabulary' in special lessons during normal school time. 'Academic vocabulary' means words that are very important for understanding school subjects, but which we don't often hear in our day-to-day conversations. For each word, children will watch a short video that explains its meaning. After watching each video, they will have an opportunity to discuss the word and do a short writing activity to help them remember it. Half of the children will watch videos in English. The other half of the children will watch videos in their home languages. There will be eight lessons like this.

At the end of the eight lessons we will test how well the children have learned the words. We will compare the scores in each group to see if hearing the explanations in home languages makes a difference to how well children learn the words.

We will also ask children to complete a questionnaire at the beginning and the end of the project. The questionnaire will ask them about the different languages they use and how they feel about them. At the end of the project, a small number of children will be invited to tell us about their experience of the project. These will be in small 'focus groups' of four or five children. The children in the group will talk with a researcher and each other about the project. These discussions will be audio recorded. We will write to you again to ask your permission if we want to invite your child to take part in these discussions.

The lessons, assessments and focus group discussions will be run by Hamish Chalmers, (see the section 'Who is conducting this research?' for information about Hamish).

## When will the project take place?

The project will take place in the summer term this year. We will work very closely with your child's teachers to ensure that the project has minimal impact on your child's normal schooling. In particular, lessons will be timetabled so that they do not interfere with your school's end of year assessments.

## What are the benefits to my child and my child's school?

By participating in this project your child and their school will be helping us to better understand ways to help multilingual children like them to do well at school. In addition, they will learn important academic vocabulary that will help them at school. We expect all children taking part in this project to develop a better understanding of these important words, no matter which group they are in.

## Are there any risks to my child and my child's school?

As far as we are aware, there are no risks involved in participating in the project. Taking part should be enjoyable and interesting for all involved.

In projects like these, sometimes people are disappointed when they are told which group they are in. This might be because they like the sound of one of the activities more than the other, for example. After we have analysed the results of the project, we will make all the videos and other materials available to the people who took part. When we do this, we will also explain what we learned by doing the project. This will help you to decide if you would like your child to try at home the version of the activity that they did not get to try when they were in the project.

## Does my child have to take part?

No. Participation in the project is entirely voluntary. Also, if you do consent for your child to be involved, but change your mind, you can withdraw that consent at any time before the results are analysed. You do not have to give a reason if you choose to do this. Your decision about whether to allow your child to take part will not have any effect on other parts of your child's schooling, such as grades and school reports.

## Will my child's privacy be protected?

Yes. All children and schools will be given numbers. These will be used instead of their names when we analyse and write about the information we collect in the study. It will then not be possible for anyone other than the researchers to work out from the number which person or school that number relates to. We promise not to share the names of the children who take part in the project with anyone else.

All data will be kept in an encrypted, password protected file on a single laptop, backed up to Oxford Brookes' University's secure online data storage system. Paper versions of the data will be kept securely in a locked cabinet in the principle investigator's office.

Oxford Brookes University requires that all data is kept for a minimum of 10 years after a project has finished. The data may be used in other projects in the future. Your child's privacy will be protected all the time.

## What will happen to the information from the study?

The study will be written up as part of Hamish Chalmers' doctoral thesis for Oxford Brookes University. A report of the study will be submitted for publication to peer reviewed academic journals and may form the basis for conference presentations. A summary of the findings will be sent to anyone involved in the project who would like one. There will be a presentation describing the main findings for parents, teachers and children involved in the project.

## How can I find out more about this project?

You are welcome to email or call Hamish if you have any questions about the project. Also, if you wish your child to stop taking part in the study later you can email him or one of the project supervisors to tell them. If you have any concerns about the way this project is being run, you can contact the Chair of the University Research Ethics Committee at Oxford Brookes University. Contact details for all Oxford Brookes University staff involved in the project are below.

Thank you for taking the time to read this leaflet. If both you and your child are happy for your child to take part in this project, please complete the attached consent form and return it to your child's class teacher.

## Contact Details

### **Principal Investigator**

Hamish Chalmers: hamish.chalmers-2015@brookes.ac.uk [REDACTED]

### **Project Supervisors**

Dr Androula Yiakoumetti: ayiakoumetti@brookes.ac.uk [REDACTED]

### **Ethics**

Chair of Oxford Brookes' Research Ethics Committee: ethics@brookes.ac.uk



# VOCABULARY LEARNING PROJECT

You are being invited to take part in a research project that will help us understand what happens when bilingual children (children who know more than one language) use both of their languages when they learn English words.

## Who is doing this research project?

The project is being run by researchers at Oxford Brookes University.

## Why am I being asked to take part in the project?

The researchers are inviting all children in Years 4, 5 and 6 at your school who know more than one language to take part.

## What is the project about and why is it being done?

The project is about what happens when children who can use more than one language are encouraged to use their non-English languages at school. We are doing this research because we are not sure whether encouraging children to use their non-English languages helps them or not. Some people think it does and some people think it doesn't. We also want to find out how children feel about using their non-English language at school.

## What will happen in the project?

We will divide the children who take part in the project into two groups. We will do this in a fair way. Both groups will be taught important words from their school subjects. One group will be taught using only English. The other group will use their other languages for some parts of the activity.

Afterwards we will find out how well children in each group know the words they were taught. We will also ask the children to complete a questionnaire at the beginning and end of the project to tell us what they thought about it.

After the project you might be invited to talk in a small group, while a researcher listens, about what it was like to take part in the project.

## Where will the project take place?

The project will take place at your school.

## How long will the project take?

There will be a total of eight lessons, spread over four weeks in the summer term. These will be timetabled so that they don't interfere too much with your normal lessons. There will also be one lesson at the beginning and one at the end of the project to find out how well you know the words, and to ask you how you feel about using your non-English language.

## What will happen after the end of the project?

After we have finished the project, we will write about it in special magazines and books for researchers and teachers. We will also tell people about the research in special talks.

## Will you use my name when you write and talk about this project?

No. We promise not to use your names or the name of your school when we write and talk about this project. We will also make sure that all the information about you and your school cannot be seen by anyone other than the researchers. We will do this by keeping information about you in a hard drive that needs a password to open it and keeping paper copies in a locked cabinet. We have to keep information about our project at the university for 10 years, but we will make sure that your names are not included in that information.

## Do I have to take part in this project?

We think that the project will be fun and interesting, but you do not have to take part in it if you do not want to. If you decide to take part, then change your mind, that's fine. You can stop being a part of the project at any time. You won't need to tell us why you want to stop.

## Which adults will work with me in this project?

Hamish Chalmers is in charge of the project. He will be the teacher you see in school. Dr Androula Yiakoumetti and Dr Ana Souza are both researchers at Oxford Brookes University. They are helping Hamish. If you have any questions about the project you can ask your parents to email these people to ask them for you. Your parents have been given their email addresses.

## What do I do now?

Talk to your parents about the project. If you are both happy for you to take part, they will give you a form to give to your teacher saying that you want to be involved in the project.

## Appendix E: Parental consent form

### CONSENT FORM

#### Vocabulary Learning Project: What are the effects of encouraging multilingual children to use their home languages to learn new words?

Hamish Chalmers, Doctoral Researcher, School of Education, Oxford Brookes University, Harcourt Hill Campus, Oxford OX2 9AT

Please initial box

1. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.
  
2. I understand that my child's participation is voluntary and that I am free to withdraw his or her participation at any time, without having to give a reason.
  
3. I agree to my child taking part in the above research project.
  
5. I agree that my child's data gathered in this study may be stored (after it has been anonymised) in a specialist data centre at Oxford Brookes University and may be used for future research.
  
6. I have discussed this project with my child.

---

Name of Child

---

Age

---

Class

---

Name of Parent or Guardian

---

Date

---

Signature

Hamish Chalmers

---

Name of Researcher

---

Date

---

Signature

## Appendix F: Language background questionnaire

You have been invited to do this questionnaire because you are able to use another language as well as English, even if it's only a little bit. This questionnaire should take about 10 minutes.

We are going to use the term 'home language' to mean the language that you use that is not English. Some people call this your 'mother tongue' or your 'first language'.

Your answers will help us to understand the way you use your home language. This will help your teachers to teach you better.

Your answers will be seen only by the researchers who are working with you at your school.

We will not tell anyone else your name or what your individual answers were.

Please answer as many questions as you can. You don't have to answer any question that you don't want to answer.

You can ask an adult to help you if you would like to.

### About You and Your Home Languages

What is your name?

What is the name of your school?

Which class are you in?

What languages are spoken in your home?

Write down all the languages used in your home. Include English, if you use it at home.

### Your Home Language

What is your home language called?

### Instructions

You click in the circle that you think is best for your answer to each question. For example, for the first question, if you cannot speak your home language at all, click on Circle 1. If you can speak it a tiny bit, tick Circle 2. If you can speak it quite well, tick Circle 3. If you can speak it very well, tick Circle 4. And if you can speak it perfectly, tick Circle 5.

1. How well do you SPEAK your home language?

I cannot speak it at all 😞 1 2 3 4 5 I speak it perfectly 😊

2. How well do you understand your home language when you HEAR it?

I do not understand it at all 😞 1 2 3 4 5 I understand it perfectly 😊

3. How good are you at WRITING in your home language?

I cannot write in my home language at all 😞 1 2 3 4 5 I am very good at writing in my home language 😊

4. How good are you at READING in your home language?

I cannot read in my home language at all 😞 1 2 3 4 5 I am very good at reading in my home language 😊

## Learning your home language

Do you have special lessons to teach you how to read and write in your home language?

Yes No

If yes:

Who teaches you how to read and write in your home language?

A member of my family

A teacher who comes to my house

I go to a special school to learn

Other

## School in another country

Have you ever been to school in another country where your home language is used?

Yes No

## School in England

How long have you been in England?

I was born in England

I came to England before I started school

I came to England in Reception or EY

I came to England in Year 1

I came to England in Year 2

I came to England in Year 3

I came to England in Year 4

I came to England in Year 5

I came to England in Year 6







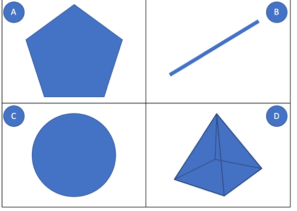
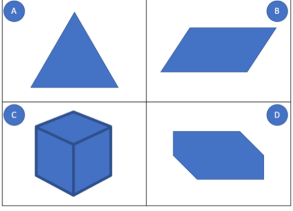
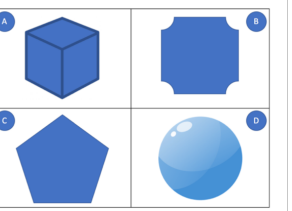

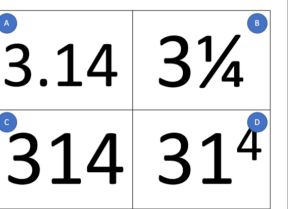
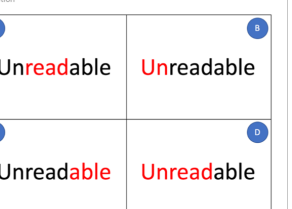
## Appendix G: Expressive vocabulary test

1. Animals that are cold blooded, have scales and lay eggs are called \_\_\_\_\_. Lizards, snakes, tortoises and crocodiles are all this kind of animal.
2. Animals that live half in water and half on land are called \_\_\_\_\_. Frogs, tadpoles, toads, and salamanders are all this kind of animal.
3. Animals that are warm blooded, have hair, and feed their babies milk are called \_\_\_\_\_. Cows, humans, dogs, cats, and dolphins are all this kind of animal
4. Animals that hunt and kill other animals for food are called \_\_\_\_\_. Eagles, ladybirds and sharks are this kind of animal.
5. Animals that are hunted, killed and eaten by other animals are called \_\_\_\_\_. Deer, rabbits, and sheep are this kind of animal.
6. Animals that eat only meat are called \_\_\_\_\_. Sharks, lions, and T-rexes are this kind of animal.
7. A 2D shape with more than two straight sides is called a \_\_\_\_\_. Squares, triangles, pentagons and hexagons are all this kind of shape.
8. A 2D shape with four sides and four corners is called a \_\_\_\_\_. Squares, rectangles, rhombuses and trapeziums are all this kind of shape.
9. A 3D shape with flat faces is called a \_\_\_\_\_. Cubes, cuboids and pyramids are all this kind of shape.
10. The name we use in Maths for a whole number is \_\_\_\_\_. Positive whole numbers, negative whole numbers and zero are all called this.
11. A number that is made up of a whole and a fraction is called a \_\_\_\_\_.  $1\frac{1}{2}$ ,  $4\frac{3}{4}$  and  $109\frac{1}{4}$  are all this kind of number.
12. A group of letters that is added to the end of a word to make it a new word is called a \_\_\_\_\_. The 'ful' in 'wonderful', the 'ness' in 'happiness', and the 'ly' in 'happily' are all called this.
13. A group of letters added to the beginning of a word to make it a new word is called a \_\_\_\_\_. The 're' in 'review', the 'non' in 'nonsense', and the 'un' in 'unusual' are all called this.
14. A word that describes a noun is called an \_\_\_\_\_. Happy, pink, warm, and noisy are all this kind of word.
15. A word that describes how something is done is called a \_\_\_\_\_. Fast, slowly, and angrily are all this type of word.
16. A group of people who come to set up a new home in a new place are called \_\_\_\_\_. The pilgrims who went to the USA were these kinds of people.
17. People who travel to different parts of the world to live, work or be with their family are called \_\_\_\_\_. Sometimes these people go to the new place for a short time, sometimes they go for a long time.
18. A place where people live is called a \_\_\_\_\_. Villages, towns, and cities are all this kind of place.

19. In geography the features of a place are either natural or made by people. If they are made by people, we call them \_\_\_\_\_ features. Bridges, roads, hospitals and schools are all this type of feature.

20. In geography the features of a place are either natural or made by people. If they are natural, we call them \_\_\_\_\_ features. Hills, rivers, and oceans are all this type of feature.

# Appendix H: Receptive vocabulary test

<p>1. Which of these pictures shows a reptile?</p> <p>Description (optional)</p> <p>Question</p> 	<p>2. Which of these pictures shows an amphibian?</p> <p>Description (optional)</p> <p>Question</p> 	<p>3. Which of these pictures shows a mammal?</p> <p>Description (optional)</p> <p>Question</p> 
<p>4. Which of these pictures shows a predator?</p> <p>Description (optional)</p> <p>Question</p> 	<p>5. Which of these pictures shows a prey animal?</p> <p>Description (optional)</p> <p>Question</p> 	<p>6. Which of these pictures shows a carnivore?</p> <p>Description (optional)</p> <p>Question</p> 
<p>7. Which of these pictures shows a polygon?</p> <p>Description (optional)</p> <p>Question</p> 	<p>8. Which of these pictures shows a quadrilateral?</p> <p>Description (optional)</p> <p>Question</p> 	<p>9. Which of these pictures shows a polyhedron?</p> <p>Description (optional)</p> <p>Question</p> 
<p>10. Which of these pictures shows an integer?</p> <p>Description (optional)</p> <p>Question</p> 	<p>11. Which of these pictures shows a mixed number?</p> <p>Description (optional)</p> <p>Question</p> 	<p>12. In which of these pictures is the suffix coloured red?</p> <p>Description (optional)</p> <p>Question</p> 



13. In which of these pictures is the prefix coloured red?

Description (optional)

Question

<b>A</b>	Uncomfortable	<b>B</b>	Uncomfortable
<b>C</b>	Uncomfortable	<b>D</b>	Uncomfortable

14. Which of the words in the sentence below is an adjective?

Description (optional)

Question

The cat played with the yellow ball.

15. Which of the words in the sentence below is an adverb?

Description (optional)

Question

The dog ate its food quickly.

16. Which of these pictures shows settlers?

Description (optional)

Question

17. Which of these pictures shows migrants?

Description (optional)

Question

18. Which of these pictures shows a settlement?

Description (optional)

Question

19. Which of these pictures shows a human feature of a place?

Description (optional)

Question

20. Which of these pictures shows a physical feature of a place?

Description (optional)

Question

Name \_\_\_\_\_

School \_\_\_\_\_

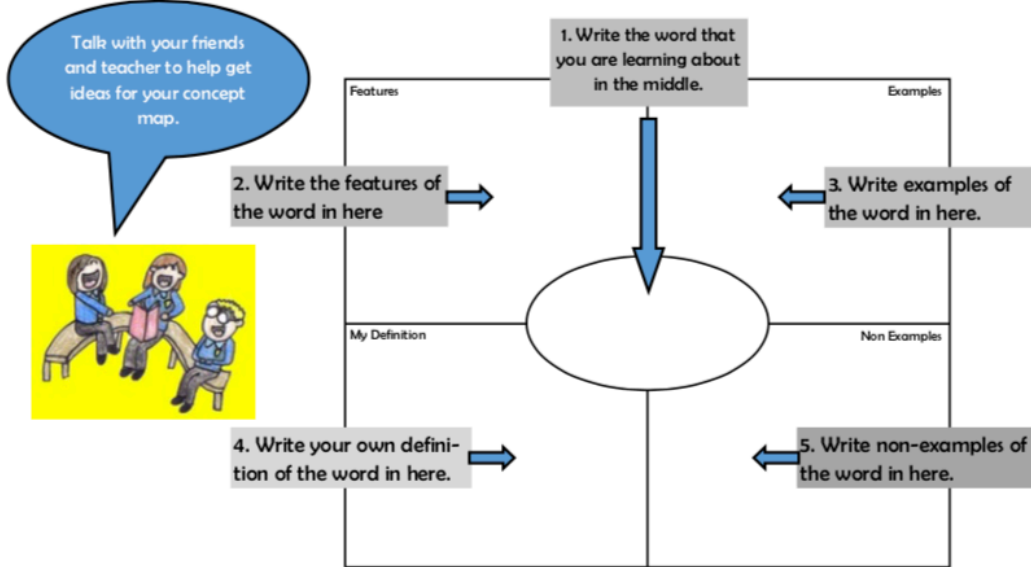
# VOGABULARY LEARNING Project



## This is a **concept map**

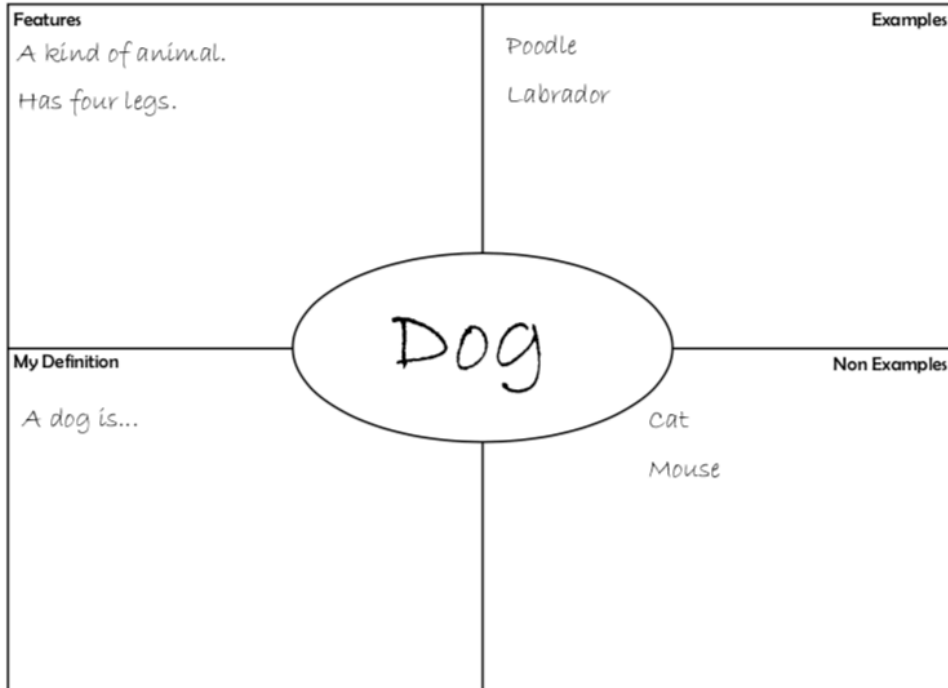
Concept maps help us to learn what words mean.

This is how you use a concept map:

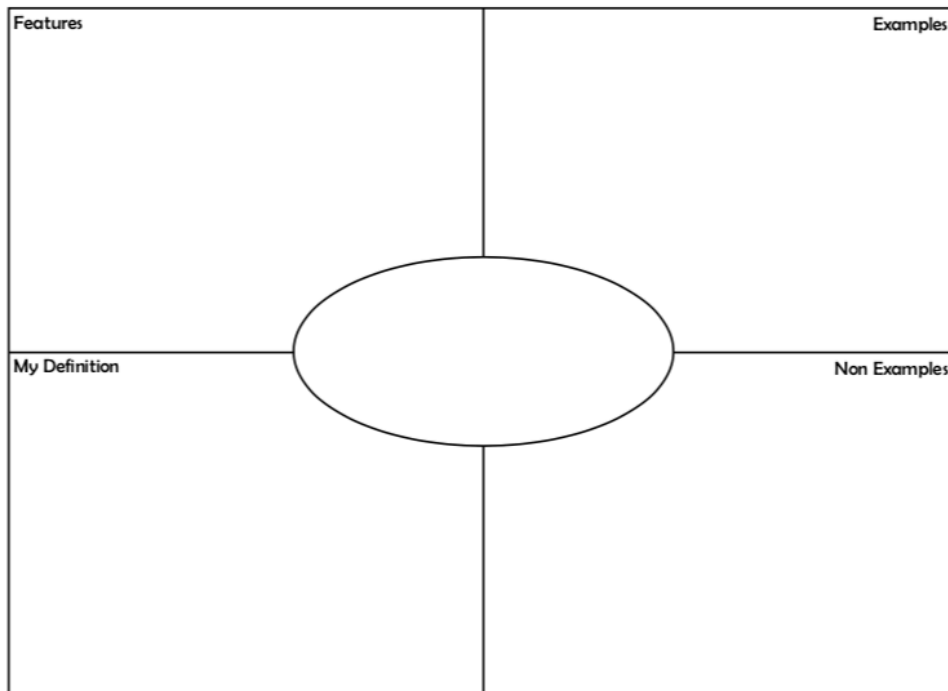


Let's do one together.

I have started each section for you. Let's talk about other things that we could write.



Date \_\_\_\_\_



N.B. From this page on, the contents of the exercise books consisted of more blank Frayer models like this one.

**Appendix J: Target items of vocabulary: Long list, Reduced list, Final items**

Original List of Target Vocabulary Items (Long List)	Reduced List of Target Vocabulary Items. Words correctly recalled or recognised in the field test by $\geq$ 80% of testers (Short List)	Final List of Target Vocabulary Items. Top two scoring words in the baseline assessment of participants removed.
Reptile	Reptile	Reptile
Amphibian	Amphibian	Amphibian
Mammal	Mammal	
Predator	Predator	Predator
Prey	Prey	Prey
Herbivore		
Carnivore	Carnivore	Carnivore
Omnivore		
Solid		
Liquid		
Gas		
Polygon	Polygon	Polygon
Quadrilateral	Quadrilateral	Quadrilateral
Polyhedron	Polyhedron	Polyhedron
Integer	Integer	Integer
Fraction		
Mixed Number	Mixed Number	Mixed Number
Suffix	Suffix	Suffix
Prefix	Prefix	Prefix
Noun		
Adjective	Adjective	
Verb		
Adverb	Adverb	Adverb
Punctuation Mark		
Invader		
Settler	Settler	Settler
Migrant	Migrant	Migrant

---

<b>Settlement</b>	Settlement	Settlement
<b>Human Features</b>	Human Features	Human Features
<b>Physical Features</b>	Physical Features	Physical Features

---