

Developing fluency with multi-word expressions

by

Haidee Elizabeth Thomson

A thesis submitted to

Te Herenga Waka

Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy in Applied Linguistics

Te Herenga Waka

Victoria University of Wellington

2020

Abstract

This three-part study was motivated by the need for empirically tested methods for teaching and learning multi-word expressions to develop fluency in language learning classrooms. Using an action research paradigm in an EFL university learning context in rural Japan, the study draws on earlier work by Boers, Eyckmans, Kappel, Stengers, & Demecheleer (2006) who linked speaking fluency with the use of multi-word expressions, and Wood (2009) who found increases in fluency after multi-word expression focussed teaching and practice with one learner in an ESL context. This study also draws on Nation's (2007) Four Strands framework for fluency building.

In the first of the three studies, a conceptual replication of the fluency workshop (including phrase instruction, shadowing, dictogloss, and role-play for example) by Wood (2009) was carried out. In contrast to Wood's approach, this study contained more than one participant ($n = 52$) and a control group ($n = 35$). The control group also followed a fluency building program but without a focus on learning target expressions. Learning effects were tested using pre- and post-test measures, including a cloze test of 30 target multi-word expressions to measure form and meaning knowledge, a dialogue role-play recording between participants to measure speaking fluency and use of multi-word expressions, and the first three levels of the Listening Vocabulary Levels Test (McLean et al., 2015) to measure general vocabulary knowledge. Feedback was collected from participants and teacher/researcher observations were recorded to evaluate the contextual appropriateness of experimental classroom activities. Results showed development of meaning and form knowledge of target expressions, but no discernible development in spoken use of the expressions or fluency for the experimental group when compared with the control group. There was also no difference in general vocabulary knowledge between the groups.

The second study ($n = 25$) incorporated improvements to the class activities, such as adding time limits to activities, and data collection methods, including expanding the role-play scenario, based on feedback from the first study. A qualitative analysis of two speakers' use of a target expression in conversation suggested that fluent use of target expressions could be achieved within nine class hours if participants had some prior knowledge of the expressions. However, a replication was necessary with more

participants and a control group to be able to generalise target expression use results to a wider population. Therefore, the third study replicated the second with more participants ($n = 65$) and a control group ($n = 51$). Results confirmed that the adapted experimental teaching activities were effective for developing form and meaning knowledge as well as the use of the target expressions in conversation. However, while fluency improved within the experimental group, the improvement was not statistically significant when compared with the control group.

The results from these studies suggest that a focus on multi-word expressions with speaking practice is helpful for developing knowledge of meaning, form and use. Therefore, language teachers are encouraged to layer up opportunities through a range of activities such as shadowing and role-play for learners to encounter and re-use frequent and useful multi-word expressions. Teachers are also encouraged to engage in action research so that they can discover learner preferences applicable to their contexts and adapt their activities to be more effective for learning and more enjoyable for learners. Developing fluency in an EFL context may indeed require more than nine intervention hours, therefore a conceptual replication of the third study with a longer intervention period would help to expand our understanding for how long it takes to attain fluency benefits from multi-word expression focused interventions in an EFL context.

Acknowledgements

I am incredibly grateful for the opportunities and support I have had in life that have led me to be able to pursue this project. I would like to acknowledge all my participants whose cooperation have made this research possible. I would like to express my gratitude and appreciation to my supervisors. I am particularly grateful to Dr. Averil Coxhead who has kept me accountable and on task throughout the research journey. Her support has motivated me to stay focused over the last five years. I am also very grateful to my secondary supervisor Dr. Frank Boers who has been a reliable critic and cheerleader throughout the years. Great appreciation to Dr. Lisa Woods for support with my statistical planning, analysis and reporting. Many thanks to Dr. Deborah Laurs for her professional writing feedback which has helped me to improve the clarity of my writing. I have appreciated the professional warmth and support of faculty and fellow post-graduates at LALS VUW, it has been a privilege to participate in such an active academic community. The morning teas and Friday seminars were always welcome breaks to look forward to.

My sincere gratitude to my work colleagues in Hokkaido, Japan, both at Muroran Institute of Technology and Hokusei Gakuen University. I am especially grateful to Eric Hagley for graciously allowing me to upload massive audio files to his server! Ngā mihi nui to Matthew Cotter for courageously picking up my JALT Hokkaido responsibilities plus some when I needed to step back to focus on writing. Thank you to my ever supportive JALT Hokkaido colleagues who have encouraged me to research effective pedagogical methods for the classroom.

I would like to thank Dr. David Wood for the inspiration to investigate these classroom intervention methods for building fluency, and for fielding my questions regarding his 2009 study. My sincerest thanks to the many people who have given me feedback, criticism or encouragement over the last 5 years.

Thank you especially to my mother (Karen Rich) for always welcoming me home for research visits. Finally, I must acknowledge my ever-positive life partner Dr. Rob Thomson for supporting my efforts and making sure my work time was all the more

efficient by enticing me to take time out to explore the outdoors with cycle touring, camping, hiking, back-country skiing and most recently canoeing.

Dedication

I would like to dedicate this thesis to my father (Michael Rich) (1950-2016) and my grandparents (Rowland Rout 1928-2018 and Elizabeth Rout 1933-2018), who saw me start this project but sadly passed away before seeing its completion.

In addition, I dedicate this thesis to all language learners and teachers who are pursuing greater knowledge and ability with multi-word expressions, may the findings and discussion within these pages assist you in your journey.

Table of contents

Abstract	i
Acknowledgements	iii
Dedication.....	iv
Table of contents	v
List of tables	xv
List of figures	xix
List of appendices	xxi
List of abbreviations.....	xxiii
 1 Chapter One: Introduction.....	 1
1.1 Why investigate fluency development?	1
1.1.1 The EFL context	1
1.1.2 How does fluency develop?	2
1.1.3 Multi-word expressions are ubiquitous in fluent speech	2
1.1.4 Helping learners develop fluency with multi-word expressions.....	3
1.2 Research outline and aims	4
1.3 Contribution of the present research	5
1.4 Organisation of the thesis.....	5
 2 Chapter Two: Literature Review	 9
2.1 The English as a foreign language learning environment.....	9

2.2 Fluency.....	11
2.3 Skill acquisition/learning theory and chunking	12
2.3.1 Concept of chunking (cognitive units).....	14
2.3.2 Frequency effects	15
2.3.3 Evidence of frequency effects from quasi-experimental studies	16
2.4 Defining multi-word expressions	17
2.4.1 Associations between use of multi-word expressions and fluency.....	19
2.5 Ways to learn multi-word expressions.....	20
2.5.1 Study abroad	21
2.5.2 Noticing.....	21
2.5.3 Learning fluency strategies with lexical fillers	23
2.5.4 Retrieval/recall.....	24
2.5.5 Memorization	25
2.5.6 Rehearsal.....	26
2.5.7 The Four Strands balance.....	27
2.5.8 Integrated skills (linked skills) programs.....	29
2.6 Explicit teaching and measurement of multi-word expressions	30
2.6.1 Limitations of the fluency workshop	32
2.7 Need for replication	34
2.8 Overall research questions	36
3 Chapter Three: Research design and overall methodology	39
3.1 Differences between Wood (2009) and this study	39
3.1.1 Case study vs. quasi-experimental study (including control group).....	40
3.1.2 ESL in Canada vs. EFL in Japan	40
3.1.3. Higher proficiency vs. lower proficiency	41
3.1.4 Monologue vs. dialogue.....	41
3.1.5 Phrase instruction using English and L1	41

3.1.6 Shadowing frequency eight times vs. four times	42
3.1.7 Traditional dictogloss vs. dictogloss with MWEs printed	42
3.1.8 Decreasing time role-play 2/1.5/1	42
3.1.9 Data collection instruments.....	42
3.1.10 Feedback on activities from participants	43
3.2 Cyclical action research approach	43
3.3 Experimental and control groups.....	45
3.4 Ethics.....	46
3.5 Experimental classroom materials	47
3.5.1 Target multi-word expressions (MWEs)	49
3.5.2 Checking the usefulness of the multi-word expressions (MWEs).....	50
3.6 Data collection instruments.....	52
3.6.1 Language learning background survey	53
3.6.2 General vocabulary knowledge.....	54
3.6.3 Dialogue recordings (measuring of fluency)	55
3.6.4 Multi-word expression cloze test	56
3.6.5 Activity feedback surveys.....	57
3.6.6 Activity feedback focus group interviews	57
3.7 Procedures	58
3.7.1 Informed consent and language learning background survey.....	58
3.7.2 Listening Vocabulary Levels Test	58
3.7.3 Dialogue recordings (measuring of fluency)	59
3.7.4 Multi-word expressions cloze test	59
3.7.5 Classroom activities for the experimental condition	60
3.7.5.1 Listen with gist questions.....	61
3.7.5.2 Marking pauses	61
3.7.5.3 Phrase instruction	61
3.7.5.4 Shadowing.....	61

3.7.5.5 Dictogloss.....	62
3.7.5.6 Mingle jigsaw.....	62
3.7.5.7 Role-play train.....	63
3.7.5.8 Decreasing time role-play	63
3.7.5.9 Record role-play	64
3.7.5.10 Free role-play	64
3.7.6 Classroom activities for the control condition	64
3.7.7 Activity feedback surveys.....	65
3.7.8 Activity feedback focus group interviews	66
3.8 Data analysis.....	66
3.8.1 The language learning background survey	67
3.8.2 Listening Vocabulary Levels Test (LVLT)	67
3.8.3 Measuring fluency in dialogue recordings.....	67
3.8.3.1 Transcription	69
3.8.3.2 Analysing the transcripts.....	70
3.8.3.3 How many multi-word expressions should we expect?.....	70
3.8.3.4 Counting use of multi-word expressions	71
3.8.4 Multi-word expressions cloze test	71
3.8.5 Activity feedback surveys.....	72
3.8.6 Activity feedback focus group interviews	72
3.9 Summary of overall methodology	73
4 Chapter Four: Study One	75
4.1 Research questions.....	75
4.1.1 Participants.....	75
4.2 Knowledge of target multi-word expressions.....	77
4.2.1 Multi-word expression cloze test gain scores	77
4.2.2 Multi-word expression cloze test within group comparisons	77
4.2.3 Answering RQ. 1	78

4.3 Multi-word expression use	78
4.3.1 Multi-word expression use gain scores	79
4.3.2 Comparing MWE use with Wood (2009)	80
4.3.3 Comparing length of multi-word expressions.....	81
4.3.4 Multi-word expression length gain scores	82
4.3.5 Comparing multi-word expression length with Wood (2009)	82
4.3.6 Extended multi-word expressions	82
4.3.7 Difference in length between MWEs and EMWEs	83
4.3.8 Length of extended multi-word expressions	83
4.3.9 Answering RQ. 2	84
4.4 Spoken fluency	85
4.4.1 Speech rate gain scores	86
4.4.2 Speech rate compared with Wood (2009).....	87
4.4.3 Mean length of runs	88
4.4.4 Mean length of runs pre- to post-measure gain scores	89
4.4.5 Answering RQ3	90
4.5 Relationship between fluency and use of multi-word expressions	91
4.5.1 Correlation between speech rate and multi-word expression use	91
4.5.2 Correlation between gains in speech rate and gains in MWE length	93
4.5.3 Correlation between gains in mean length of runs and MWE use.....	93
4.5.4 Answering RQ. 4	94
4.6 Aural vocabulary knowledge	95
4.6.1 Listening Vocabulary Levels Test (LVLT)	95
4.6.2 Pre-test vocabulary scores.....	95
4.6.3 Post-test vocabulary scores	96
4.6.4 Vocabulary gain scores	96
4.6.5 Vocabulary scores by level	96
4.6.6 Answering RQ 5	97
4.7 Participant feedback regarding experimental activities	97

4.7.1 Marking pauses	98
4.7.2 Role-play recording	100
4.7.3 Dictogloss	101
4.7.4 Mingle jigsaw.....	103
4.7.5 Decreasing time role-play	105
4.7.6 Free role-play	108
4.7.7 Shadowing.....	109
4.7.8 Role-play train	110
4.7.9 Phrase instruction.....	112
4.7.10 Answering RQ. 6	114
4.8 Study One results summary	114
5 Chapter Five: Study Two method and results	115
5.1 Changes to improve data collection instruments	115
5.1.1 Expand scenario for pre- and post-intervention dialogue recording.....	115
5.1.2 Adapt activity feedback survey	116
5.1.3 Change from five-point Likert scale to six-point Likert scale.....	117
5.1.4 Add feedback for each thematic unit	117
5.2 Changes to improve data collection methods	118
5.2.1 Familiarise all participants with dialogue recording task	118
5.2.2 Individual speakers self-identify at start of recordings.....	119
5.2.3 Instruct participants to record pre-intervention scores.....	119
5.2.4 Change in focus group recruitment method	119
5.2.5 Dialogue recording: From MP3 to WAV format.....	120
5.2.6 PRAAT analysis details	120
5.3 Changes to improve classroom activities	122
5.3.1 Remove marking pauses activity	123
5.3.2 Provide structure and class time for reflection on role-play recording.....	123
5.3.3 Add scaffolding to dictogloss hand-out	124

5.3.4 Set five-minute time limit for mingle jigsaw	124
5.3.5 Shorten hotel model dialogue	125
5.3.6 Display target multi-word expressions	125
5.3.7 Shorten initial timings for decreasing time role-plays	126
5.3.8 Reduce choice in free role-play	126
5.4 Study Two	127
5.4.1 Research questions	127
5.4.2 Study Two participants	127
5.4.3 Measuring effect	128
5.5 Multi-word expression cloze-test gain scores	129
5.5.1 Multi-word expression cloze-test gain scores compared with Study One	129
5.6 Use of multi-word expressions in conversation	130
5.6.1 Speech analysis of eight participants	130
5.6.2 Comparing multi-word expression use with the first study	131
5.6.3 Comparing range of multi-word expression use with the first study	131
5.6.4 Individual use of target multi-word expressions in conversation	132
5.7 Spoken fluency	133
5.7.1 Speech rate compared with Study One experimental group	133
5.7.2 Individual speech rate change for Study Two	134
Extract 1 Speaker A2	136
Extract 2 Speaker D1	137
5.7.3 Phonation time ratio	137
5.7.4 Articulation rate	138
5.7.5 Mean length of runs compared with Study One	139
5.7.6 Individual change in mean length of runs for Study Two participants	140
5.8 Aural vocabulary knowledge	142
5.8.1 Aural vocabulary knowledge compared with Study One	142
5.8.2 Aural vocabulary knowledge for Study Two alone	143

5.9 Which activities are considered the most useful by learners?	144
5.9.1 Everyone agreed that shadowing was useful	145
5.9.2 Role-play activity feedback results	146
5.9.3 Phrase instruction activity feedback results	146
5.9.4 Mingle jigsaw activity feedback results	146
5.9.5 Decreasing time role-play activity feedback results	147
5.9.6 Dictogloss activity feedback results	148
5.9.7 Free role-play activity feedback results	148
5.9.8 Recording the role-play activity feedback results	149
5.9.9 Unit themes useful and enjoyable	149
5.10 Study Two results summary	150
6 Chapter Six: Study Three method and results	153
6.1 Changes made to improve classroom activities	153
6.1.1 Shadowing: Add explicit instruction to allow pausing	153
6.1.2 Hotel unit decreasing time role-play: Provide booking form realia	154
6.1.3 Hotel unit decreasing time role-play: Increase time	154
6.2 Study Three	154
6.2.1 Research questions	155
6.2.2 Study Three participants	155
6.2.3 Measuring effect	157
6.3 Multi-word expression cloze test results	157
6.3.1 Replicable effect	158
6.4 Use of multi-word expressions in conversation	159
6.4.1 Between group comparisons of multi-word expression use	159
6.4.2 Comparing multi-word expression use with previous studies (by dyad) ..	160
6.4.3 Comparing range of multi-word expression use	161
6.5 Spoken fluency	163

6.5.1 Speech rate	163
6.5.1.1 Comparison of gains in speech rate across the three studies (by dyad)	164
6.5.1.2 Correlation between speech rate and multi-word expression use ...	164
6.5.2 Phonation time ratio	165
6.5.3 Articulation rate	167
6.5.4 Mean length of runs (untrimmed)	168
6.5.4.1 Mean length of runs (by dyad) compared with previous studies	169
6.5.4.2 Correlation between mean length of runs and MWE use	170
6.5.5 Summary regarding fluency outcomes	171
6.6 Aural vocabulary knowledge	172
6.7 Which activities were considered most useful by the learners?	175
6.8 Study Three results summary	177
7 Chapter Seven: Discussion.....	179
7.1 Learning from replication of the fluency workshop.....	180
7.1.1 How action research can inform activities in the classroom.....	181
7.2 Stages of skill development for chunking multi-word expressions	182
7.2.1 Cognitive/declarative/presentation stage measured by cloze test.....	183
7.2.2 Chunking and connection with vocabulary knowledge	184
7.2.3 Associative/procedural stage (multi-word expression use)	185
7.2.4 Number of encounters	185
7.3 Automatic stage: Relationship between multi-word expression use and fluency.....	187
7.3.1 Building fluency in an EFL context: Nine hours is not enough	188
7.4 Four Strands and linked skills in a language course	190
7.5 Conclusion	192

8	Chapter Eight: Implications and conclusion	195
8.1	Summative claims and theoretical implications.....	195
8.2	Pedagogical implications	197
8.2.1	Selection of multi-word expressions.....	197
8.2.2	Activity sequencing for fluency	198
8.2.3	Action research and learner feedback for improved classroom learning..	200
8.3	Limitations and implications for methodology	203
8.3.1	Six-week time frame	203
8.3.2	Difficulties gathering data from learners	204
8.3.3	Comparability between research data sets	205
8.3.4	Difficulty measuring procedural knowledge	205
8.4	Future research	207
8.4.1	Extended replication opportunities	207
8.4.2	Using alternative measurement methods	208
8.4.3	Possible post-hoc analysis.....	208
8.5	Reflections on the PhD journey (what have I learnt)	209
9	References.....	211

List of tables

Table 2.1 Overview of activities in the fluency workshop	31
Table 2.2 Results from fluency workshop	32
Table 2.3 Overview of research questions for three studies	37
Table 3.1 Differences between Wood (2009) and the current study.....	40
Table 3.2 Participants proficiency range over three studies	46
Table 3.3 Word profile of model texts for Study One	47
Table 3.4 The thirty multi-word expressions targeted in the study.....	50
Table 3.5 Frequency profile of words from the multi-word expressions.....	52
Table 3.6 Fluency workshop activities in sequential order.....	60
Table 4.1 Study One participant numbers by group with TOEIC score range	76
Table 4.2 Multi-word expression use in the pre- and post-measure	78
Table 4.3 Mean length of expressions by group	83
Table 4.4 Trimmed speech rate in the pre- and post-test	85
Table 4.5 Descriptive statistics for pre- and post-measure untrimmed speech rate	86
Table 4.6 Descriptive statistics for pre- and post-measure mean length of runs ..	89
Table 4.7 Descriptive statistics for pre- and post-measure runs per minute	91
Table 4.8 Descriptive statistics for LVLTL results across groups	95
Table 4.9 Listening Vocabulary Levels Test results by level and group	97
Table 4.10 Model dialogue word count and speech rate for Study One	107
Table 5.1 Changes to improve data collection instruments	115
Table 5.2 Summary of changes to data collection methods in Study Two	118
Table 5.3 Outline of changes to classroom activities for Study Two	123

Table 5.4 Word count, time and conversation turns for model audio tracks	125
Table 5.5 Participant numbers by group with TOEIC score percentages	128
Table 5.6 Measures with purpose	129
Table 5.7 Change in number of words spoken as part of multi-word expressions	133
Table 5.8 Speech rate before and after treatment.....	135
Table 5.9 Phonation time ratio change between pre- and post-measure.....	138
Table 5.10 Articulation rate	139
Table 5.11 Mean length of runs for Study Two participants	140
Table 5.12 Change between pre- and post- mean length of runs	141
Table 5.13 LVL differences by level	143
Table 5.14 Activity usefulness as agreed by participants	145
Table 6.1 Changes made to experimental classroom activities for Study Three	153
Table 6.2 Study Three participants' TOEIC score range percentages	155
Table 6.3 Gender and major for Study Three participants.....	156
Table 6.4 Measures used in Study Three with their purposes.....	157
Table 6.5 Descriptive statistics for Study Three multi-word expression cloze test	158
Table 6.6 Descriptive statistics for Study Three multi-word expression use.....	160
Table 6.7 Descriptive statistics for Study Three multi-word expression use (by dyad)	161
Table 6.8 Post-intervention use of complete expressions	162
Table 6.9 Descriptive statistics for untrimmed speech rate	163
Table 6.10 Descriptive statistics for phonation time ratio in Study Three.....	166
Table 6.11 Descriptive statistics for articulation rate for Study Three by group	168

Table 6.12 Descriptive statistics for mean length of runs in Study Three	169
Table 6.13 Descriptive statistics for mean length of runs by dyad in Study Three	170
Table 6.14 Listening Vocabulary Levels Test results within group by level.....	174
Table 6.15 Activity usefulness as agreed by Study Three participants.....	176

List of figures

Figure 3.1 The action research cycle.....	44
Figure 3.2 Example of model dialogue for café unit	48
Figure 4.1 Average use of multi-word expressions by group	81
Figure 4.2 Speech rate in the pre- and post-measure compared with Wood (2009)	88
Figure 4.3 Correlation between change in speech rate and change in multi-word expression use.....	92
Figure 4.4 Correlation between change in speech rate and change in length of multi-word expressions	93
Figure 4.5 Mean length of run gain correlated with multi-word expression use gain	94
Figure 4.6 Activity feedback survey results for marking pauses	99
Figure 4.7 Activity feedback survey results for role-play recording	101
Figure 4.8 Activity feedback survey results for dictogloss.....	103
Figure 4.9 Activity feedback survey results for mingle jigsaw	105
Figure 4.10 Activity feedback survey results for decreasing time role-play	107
Figure 4.11 Activity feedback survey results for free role-play	109
Figure 4.12 Activity feedback survey results for shadowing.....	110
Figure 4.13 Activity feedback survey results for role-play.....	112
Figure 4.14 Activity feedback survey results for phrase instruction	113
Figure 5.1 Adjusted activity feedback survey example	117
Figure 6.1 Correlation between speech rate and multi-word expression use.....	165
Figure 6.2 Box plot for phonation time ratio gain across groups.....	167
Figure 6.3 Correlation between mean length of runs and multi-word expression use.....	171

Figure 6.4 LVL T gain score comparison between groups	173
---	-----

List of appendices

Appendix 1 Ethics approval from Japanese institution.....	228
Appendix 2 Ethics approval from New Zealand institution.....	229
Appendix 3 First amendment to ethics approval.....	230
Appendix 4 Second amendment to ethics approval	231
Appendix 5 Learner information and consent page (English translation)	232
Appendix 6 Learner information and consent page (Japanese)	234
Appendix 7 Participant information sheet and consent form for the focus group (English translation)	236
Appendix 8 Participant information sheet and consent form for the focus group (Japanese)	239
Appendix 9 Model dialogues	242
Appendix 10 The 30 multi-word expressions with frequency and function information	248
Appendix 11 Language learning survey (English translation).....	251
Appendix 12 Language learning survey in Japanese	253
Appendix 13 Pre- and post-intervention dialogue instructions for Study One...	256
Appendix 14 Café menu (realia) that participants discussed for the dialogue....	257
Appendix 15 Multi-word expression cloze test with answers and screenshot....	258
Appendix 16 Activity feedback survey format example with translation.....	262
Appendix 17 Dictogloss worksheet from the café unit in Study One.....	263
Appendix 18 List of multi-word expressions (MWEs) used in Study One	264
Appendix 19 List of extended multi-word expressions used in Study One.....	267
Appendix 20 Adapted pre- post- intervention role-play instructions.....	268
Appendix 21 Adjusted activity feedback survey example	269

Appendix 22 Technical support confidentiality agreement	270
Appendix 23 Fluency reflection worksheet	271
Appendix 24 Adjusted dictogloss example from café unit	273
Appendix 25 List of multi-word expressions (MWEs) used in Study Two.....	274
Appendix 26 Hotel booking form realia from Study Three	275
Appendix 27 List of multi-word expressions (MWEs) used in Study Three.....	276

List of abbreviations

Abbreviation	Full form
EFL	English as a foreign language
ESL	English as a second language
L1	First language
L2	Second language
SLA	Second language acquisition
MWE	Multi-word expression
LVLT	Listening Vocabulary Levels Test
TOEIC	Test of English for international communication

1 Chapter One: Introduction

In this first chapter of eight, I justify why I decided to investigate fluency development, the development of multi-word expression knowledge, and classroom interventions to improve both. I introduce the aims and the scope of the research, and how I perceive the present research contributes to wider knowledge. In the final section of this chapter, I explain the organisation of the thesis with brief outlines of the content for each of the upcoming seven chapters.

1.1 Why investigate fluency development?

There are several reasons why it is important to investigate the relationship between fluency, multi-word expressions, language teaching and learning. Firstly, fluency is difficult to achieve for adult language learners, especially those learning without an immersion environment. Secondly, if we understand how fluency develops then we can strategically approach this language learning goal. Thirdly, the ubiquitous presence of multi-word expressions in fluent speech suggests that we need to learn more about the relationship between multi-word expressions and fluency. Finally, we need to discover ways to help language learners develop fluency and it appears that helping learners to understand and use common multi-word expressions will assist in fluency development.

1.1.1 The EFL context

Achieving fluency in a new language is a challenging goal for language learners and especially difficult for learners of English as a foreign language who have fewer opportunities to use English in their daily lives. As an English language instructor in Japan for almost 10 years, I have often been asked by learners about how they can improve their speaking fluency. Fluent speakers of English are not as common as one might expect in Japan (Herder & Sholdt, 2014; Nishino & Watanabe, 2008), despite being one of the largest trading economies in the world. It is this background that shapes my focus of enquiry for this research thesis.

1.1.2 How does fluency develop?

Fluency is a “complex phenomenon that encompasses a multitude of linguistic, psycholinguistic, and sociolinguistic features” (Freed et al., 2004, p. 279). The development of automatic skill using language can be described as fluent language use. The declarative/procedural model of memory describes our overall long-term memory storage systems (not only for language). This model leads us to expect explicit and implicit learning of new knowledge and skills to be stored in declarative memory. Declarative memory refers to knowledge about something that is able to be quickly acquired but requires effort to use (Ullman, 2015). With practice, procedural memory (implicit knowledge of rules, sequences, categories, or knowledge about how to do something) also develops and finally culminates into automatic function, whereby an action can be made with little effort or thought (Ullman & Lovelett, 2018). First language (L1) learning is known to be more reliant on procedural memory which is most efficient in young learners, whereas second language (L2) learning is more reliant on declarative learning and memory (Ullman, 2015; Ullman & Lovelett, 2018). Skill acquisition theory (Anderson, 1983, 2000; DeKeyser, 2015) describes initial learning as developing cognitive/declarative knowledge, with associative/procedural knowledge developing after practice, and eventually becoming autonomous/automatized. Based on the declarative/procedural model and skill acquisition theory, we can anticipate that large amounts of practice with the L2 will be required to develop procedural knowledge to a point where L2 language use can become fast and effortless.

1.1.3 Multi-word expressions are ubiquitous in fluent speech

A large proportion of fluently spoken language is produced using multi-word expressions, which are a type of highly frequent formulaic language (Altenberg, 1998; Erman & Warren, 2000; Kuiper, 2004). Multi-word expressions are essentially short word sequences that are frequently used or chunked together. The term chunking was originally used by Miller (1956) to describe how we overcome some of the limitations on our immediate memory processing capacity by chunking bits of information together as they become familiar, and as familiarity increases so does the length of the chunks allowing us to process more information. Newell (1990) argues that chunking is an

“ubiquitous feature of human memory” (p. 7). N. C. Ellis (2001) suggests that language is learnt through making connections between chunks of language from morpheme connections to multi-word connections. We can call this the chunking theory. The more words that learners know, the easier it should be to make associations between words (chunk them together) and build knowledge of the multi-word expressions which are associated with fluent speech. There is general recognition in the research community of the processing and cognitive advantages of formulaic language (Conklin & Schmitt, 2012). Fluency benefits related to knowing and using multi-word expressions for faster comprehension or communication are more evident for first language (L1) speakers who naturally will have had greater exposure to multi-word expressions than language learners (N. C. Ellis et al., 2008; Siyanova-Chanturia et al., 2011). Therefore, it is argued that greater knowledge of multi-word expressions assists with greater fluency. As a teacher and researcher, I want to investigate the connection for learners between use of multi-word expressions and fluency. Therefore, I must also investigate teaching interventions that can help learners know and use multi-word expressions.

1.1.4 Helping learners develop fluency with multi-word expressions

Evidence of the efficacy of teaching interventions for learning multi-word expressions are not common in the literature, but there are a few that stand out. Fluency benefits from noticing and or explicit teaching of multi-word expressions have been reported in studies such as Boers et al., (2006); McGuire & Larson-Hall (2018) and Wood (2009). These studies make promising suggestions for learning interventions in the classroom that build knowledge of multi-word expressions. Wood (2009) gives a particularly detailed description of a set of classroom activities for building knowledge of multi-word expressions using activities such as shadowing, dictogloss and role-play. The studies have focused on the gain in fluency as a result of teaching interventions which is very useful, however the uptake (gain in knowledge and use) of multi-word expressions by learners has not been measured. Therefore, this study investigates specifically the uptake of 30 multi-word expressions using activities inspired by Wood (2009). The study will help evaluate effective classroom interventions that are pedagogically realistic and adaptable for varying contexts for developing fluency with multi-word expressions.

Finally, another important and not often reported aspect of learning multi-word expressions is the learner experience of such intervention. The studies mentioned above did not ask the learners how they felt about the classroom activities. Classroom interventions are experienced by the learners, therefore their enjoyment or not of activities is likely to affect how well they learn as a result. Therefore, in this study I also set out to ask and reflect upon the learner experience with the intention to discover ways to improve any activities that learners express dissatisfaction with. In the next section, I set out the overall aims and scope for this thesis.

1.2 Research outline and aims

This research thesis is my evidence-based contribution to the conversation about the best ways to teach formulaic language and the effect such teaching can have on L2 speaking fluency. I used a quasi-experimental classroom design to compare pre- and post-intervention measures between convenience sampled experimental and control groups. I have chosen to investigate the efficacy of classroom activities for learning multi-word expressions by replicating and extending the classroom activities from Wood (2009). The research by Wood (2009) will be described in greater detail in the upcoming chapters. An action research framework (Burns, 2010) was used in order to adapt and improve the original classroom activities for the specific learners and context. Three iterations with new participants for each six-week study were made with feedback from the previous iteration used to improve activities in the following study.

To address the gap in evidence for teaching interventions that increase fluency using multi-word expressions, the present research has been conducted with three primary aims. The first aim is to examine the effectiveness of explicit teaching and practice of multi-word expressions for EFL learners. Investigating the first aim requires comparing pre- and post-intervention knowledge and spoken use of multi-word expressions, along with effects on speaking fluency. The second aim is to maximise the effectiveness of classroom teaching activities through reflection on feedback from participants and researcher observations. Investigating the second aim requires collecting feedback from participants and reflecting on classroom interactions and learning results. The third aim is to verify or dispute the current claims from research

regarding 1) the relationship between learner use of multi-word expressions and fluency, 2) the skill acquisition process, and 3) the relationship between vocabulary and ability to learn multi-word expressions.

1.3 Contribution of the present research

The current research contributes to collectively held knowledge in several ways. First, it provides quasi-experimental evidence that previously was lacking for the efficacy of classroom teaching and learning activities for learning multi-word expressions in an EFL learning context. In particular, the study provides evidence of the efficacy for classroom activities similar to those in Wood (2009), known collectively as “the fluency workshop.” Second, the study reveals the preferences of Japanese learners in this study regarding interactive classroom activities. Third, the research results support previous findings and provide further evidence for the positive influence of multi-word expression use on learner speaking fluency, the skill acquisition process and the relationship that vocabulary knowledge has with learning multi-word expressions. The findings of this research therefore provide teachers, program designers, textbook writers and researchers with valuable information regarding the creation, selection and adaptation of classroom learning materials and activities. The findings also indicate how much intervention time might be required to develop learner confidence in an EFL context to use multi-word expressions in conversation.

1.4 Organisation of the thesis

This thesis consists of eight chapters. In Chapter Two, I summarise the relevant literature and current theory about multi-word expressions and fluency. I look at research regarding teaching multi-word expressions, particularly the Wood (2009) study, and argue for the importance of replication in this area of enquiry. I also introduce the action research framework (Burns, 2010) that was used to reflect upon and improve the classroom activities and research method with each iteration. The overall research questions are positioned at the end of Chapter Two with a focus on investigating the effectiveness of the explicit teaching intervention for developing knowledge and use of multi-word expressions, as well as building speaking fluency.

In Chapter Three, I describe the method and materials for Study One which was an extended replication study of Wood (2009). I then introduce the experimental intervention and control group activities. Methods for measuring pre- and post-intervention multi-word expressions knowledge, ability to use multi-word expressions, fluency measures (such as speech rate and mean length of runs) and measures of vocabulary knowledge are considered, and choices are explained. The methodology described in this chapter is the base for iterations and improvements in the second and third studies.

Chapters Four through Six report on the results of the three iterative studies, each beginning with research questions which contribute to answering the overarching research questions stated at the end of the literature review in Chapter Two. In Chapter Four, I describe and evaluate the data collection and results from the first study. Using quantitative statistical analysis, declarative knowledge of taught multi-word expressions in the experimental group was shown to increase but use of multi-word expressions and fluency were not shown to significantly increase when compared with the control group. Feedback from participants and researcher observations are quantitatively and qualitatively analysed and reported.

In Chapter Five, changes to the method and materials are described based on feedback from participants and researcher observations from Study One. The Study Two results are reported using quantitative and qualitative analysis. In Study Two, there was no control group and fewer participants, so results are difficult to apply to a wider population. However, the results are informative for analysing changes within the participants, such as theorising how long fluency with multi-word expressions takes to develop and under what conditions. Learner responses to the classroom activities were again collected, revealing that adaptations made to the classroom activities appeared to have made the activities more enjoyable for the learners.

In Chapter Six, Study Three is reported with minor adaptations to methods and materials based on participant feedback and researcher observations from Study Two. Study Three was a quasi-experimental study with more participants than the previous studies and a control group. Results and feedback are quantitatively and qualitatively

analysed and reported. Results supported the trends found in the previous studies and also revealed greater use of the multi-word expressions by the experimental group, though not greater fluency when compared with the control group.

In Chapter Seven, I discuss the overall results from the three studies under four themes. They are: (1) what we have learnt from this replication and action research, (2) stages of skill development for chunking of target multi-word expressions, (3) how fluency develops in an EFL environment, and (4) teaching fluency as part of a language course.

In Chapter Eight, I describe the theoretical, methodological and pedagogical implications from the study for classroom teaching, learning and research. I reflect on the various limitations that need to be considered when interpreting the results and how the results apply to wider populations. I also suggest future research possibilities that have emerged from this research project. I complete the chapter and thesis with a reflection on this PhD journey.

2 Chapter Two: Literature Review

This research project investigates how to teach multi-word expressions, and whether knowledge and use of such expressions can improve speaking fluency in an English as a foreign language (EFL) context. Wood's (2009) fluency workshop is replicated and refined through student feedback for the EFL context in Japan. In this chapter, I consider what the literature tells us regarding 1) the EFL environment, 2) fluency, 3) skill acquisition theory and chunking, 4) multi-word expressions and their associations with fluency, 5) teaching multi-word expressions as part of a language course, 6) explicit teaching and measurement of multi-word expressions in a fluency workshop, 7) replication, and 8) action research. The chapter culminates with research questions arising from the literature reviewed and looks toward a quasi-experimental replication of Wood (2009) using an action research methodology.

2.1 The English as a foreign language learning environment

In my EFL teaching context in northern Japan (Hokkaido), I often encounter a lack of fluency and confidence to speak English among university undergraduate students. In EFL environments around the world, for example in Saudi Arabia (Al-Nasser, 2015), the Republic of Angola (Albino, 2017), and Sri Lanka (Samaranayake, 2016), lack of confidence and fluency when it comes to speaking is a common complaint due to lack of speaking practice opportunities. English as a second language (ESL) learners can benefit from environments rich in fluency practice opportunities. The same is not true for EFL learners, who have much less opportunity outside the classroom for speaking practice. In cities or highly touristed areas in Japan, there may be more opportunities to interact with English speaking visitors or residents or see signs in English. However, in the more rural areas, these opportunities to use or see English are much less available. The internet has plenty of material in English for those who are motivated to search for it and use it for language study (Nunan & Richards, 2015). However, in my teaching context where engineering is the focus and there are only two English classes as part of the university degree program in a rural area; students are unlikely to make the effort to search out English resources online. EFL contexts often have the added tool of a shared first language (L1). L1 explanations of grammar and vocabulary are able to reduce cognitive load and fast track understanding (Bruen &

Kelly, 2017; Laufer & Girsai, 2008; Macaro, 2005). The shared use of the Japanese language in Japan can be a language learning resource. However, a focus on language form is often at the expense of practising using new language orally and developing fluency. The grammar-translation method “word by word translation of English texts into Japanese” with grammar explanation from the teacher using the shared L1 is a popular English teaching method in Japan (Morita, 2015, p. 514). Another popular language teaching method used in Japan is the audio-lingual method. Based on ideas in behavioural psychology, learners acquire language through repetition; they repeat and memorise speech in the L2. Communicative Language Teaching (CLT) is encouraged in Japan in theory. However, CLT is not used as much as the passive and highly controlled grammar-translation and audio lingual methods in the classroom due to lack of teacher training and confidence (M. Cook, 2012). English lessons are often conducted in Japanese and there is little opportunity to use English for communication. During high school, students tend to experience a focus on accuracy and test preparation at the expense of any fluency practice (Herder & Sholdt, 2014). Curricular guidelines for English language education in Japan are set at a national level by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). While these curricular guidelines emphasise the importance of developing communicative ability in English, there are constraints on implementation such as lack of teacher training and textbooks which focus on grammar and translation activities (Glasgow & Paller, 2014). There is limited time for class planning and priority is given to grammar instruction to prepare students to pass entrance exams for high schools. Since 1987, there have been foreign assistant language teachers (speakers of English) in schools, as part of the Japan Exchange and Teaching program. The presence of these assistant language teachers increases the opportunity for learners to speak and enjoy using English, however their presence is generally focused on culture mixing rather than language per se (see Nishino & Watanabe, 2008). The absence of fluency activities from the EFL classroom contributes to the lack of fluency and confidence to speak. Japan, therefore, is no exception to this wider EFL concern (Herder & Sholdt, 2014; Nishino & Watanabe, 2008).

Another factor that influences the learning environment in Japan is the cultural tendency to avoid uncertainty (Hofstede, 2011). There is a certain amount of risk involved in speaking in a foreign language, especially if you place a high value on accuracy. In order to avoid risk, a learner might decide not to speak, thereby foregoing learning opportunities. Another cultural influence in Japan is the value placed on group membership (collectivism) or fitting into the group, so it becomes difficult to speak if no one else is (Hofstede et al., 2010). Speaking ability is an important skill to develop; therefore, opportunities to speak and practice must be offered in the classroom if language learners are to gain any confidence to speak in a foreign language. Ideally such opportunities to speak will be interactive following the social cultural theory that language helps us relate to our social-material world and that we learn from each other such as when a “more knowledgeable other” helps us (Vygotsky, 1978). Activities should also be within the learners “zone of proximal development” with “scaffolding” provided to help achieve this if necessary (Vygotsky, 1978). In the EFL context, and in my context in Japan specifically, I need to identify classroom activities that can help my students develop their speaking fluency while reducing the effect of any cultural barriers to participation.

2.2 Fluency

Language fluency can be approached from many angles and definition depends on the purpose for which it is defined. Chambers (1997) states that “fluency is about effectiveness of language use within the constraints of limited linguistic knowledge” (p. 536). Speaking fluency does not necessarily mean that one is able to produce perfect flowing language in all situations. Rather, fluency refers to the ability to express oneself without undue hesitation in the situations that one desires to communicate in (Skehan, 1996). Speaking fluency in this study therefore refers to the “speed and smoothness of oral delivery” (Lennon, 1990, 2000, p. 25). Key indicators or measures of speaking fluency are: Speech rate (mean number of syllables per minute including pause time), mean length of runs (mean number of syllables between pausing), and phonation time ratio (percentage of time spent speaking) (Kormos & Dénes, 2004).

When speaking, one must balance time pressure with the construction of an easily understandable message for the listener. If the speaker is creating a message word by word, it is likely to take them longer to get their message across, pausing as they consider which word to use next. A speaker who manages to balance pre-fabricated constructions with novel constructions is likely to give the impression of a more fluent speaker (Chambers, 1997). Highly fluent speakers can speak swiftly and comprehensibly, with seemingly little effort which may be because they chunk words together as prefabricated constructions.

2.3 Skill acquisition/learning theory and chunking

This section introduces skill acquisition theory (DeKeyser, 2015) as a cognitive model for learning, followed by the concept of chunking, frequency effects, a description of multi-word expressions and finally their connection with fluency. Skill acquisition theory has its roots in psychology and has been used to understand learning in a wider range of domains from cognitive classroom learning to motor skill applications in sports and industry (DeKeyser, 2015). Using the Adaptive Control of Thought (ACT) framework for describing cognition, Anderson (1983, 2000) claimed that we develop two types of long term memory, *declarative* knowledge which is knowledge about something, and *procedural* knowledge which is knowledge learnt by doing (or in the case of language learning we might say procedural knowledge is acquired by using). The general claim is that there are three stages of knowledge and skill development.

The first stage of development is referred to as the declarative or cognitive stage (Dörnyei, 2009, p. 153). Anderson (1983, 2000) argues that all knowledge starts out in a declarative form which must be interpreted by general memory procedures. The declarative stage is argued to be very important for setting the platform for progression to the procedural and automatic stages (Dörnyei, 2009). The declarative stage in the classroom is often where the teacher helps learners to understand the form, meaning and use of a language item to be learnt through presentation, following the common teaching approach of presentation-practice-production (PPP) (Byrne, 1986) seen in teacher guidebooks such as Scrivener (2005). The learner notices the form and

comprehends the meaning of the language item. Comprehensible input in the L2 is essential for language acquisition according to the input hypothesis of the monitor theory (Krashen, 1981). Such input should be “slightly beyond the current level of the learner’s internalized language ($i + 1$)” (VanPatten & Williams, 2015, p. 26). Krashen argues that learners should also be in a stress-free comfortable state so that their *affective filters* do not hinder learning from input.

The next development stage involves turning the declarative knowledge into procedural knowledge through using the language item in communication. As a language item is practised in communicative contexts, the strength of production is said to increase with every successful attempt and “proceduralisation gradually replaces the interpretive application” (Anderson, 1983, p. 34). Towell, Hawkins, & Bazergui (1996) argue that fluent L2 production requires that knowledge become proceduralised (p. 85). Therefore, in order to produce speech at a fluent speed, the language knowledge needs to be used (practised) to develop from procedural to automated knowledge. DeKeyser (1997) reports that learning second language grammar rules with practice results in systematic reduction in reaction time and error rate. When the reaction times and error rate are charted, the most dramatic reductions in reaction time and error rate occur in the initial learning stages with a steep learning curve that flattens out as practice produces automatic/fluent activity (DeKeyser, 1997, 2015). Such a learning curve is referred to as the “power law of practice” (Dörnyei, 2009, p. 152) or the “power law of learning” (DeKeyser, 2015, p. 96). The power law of learning curve is also seen “in other cognitive domains from geometry to computer programming” (DeKeyser, 1997, p. 214). Once the learner has encountered or used the item enough times to know it rather well, a plateau stage is reached whereby extra encounters or use do not result in a whole lot more speed or accuracy. Usage based approaches to language learning see language learning as “the learning of constructions” or associations, “pairings of form and meaning or function.” (N. C. Ellis & Wulff, 2015, p. 75). The more reliable the association between the meaning and form, the easier the language item is to learn. This section has described how new language is learnt through declarative, procedural and automatic skill stages. The next section will introduce the concept of chunking and how chunking contributes to procedural skill stages in language learning.

2.3.1 Concept of chunking (cognitive units)

Bybee (2008) describes chunking as a property of procedural knowledge. Jiang (2009, p. 102) notes that “vocabulary learning per se should not be taken as remembering as many difficult words as possible, but rather learning how to combine the more frequent ones.” With limits to our processing capacity or working memory it is only natural that we attempt to streamline and shortcut what we can. People often streamline their thought processes through chunking related information or skills together. Dörnyei (2009) states that in the process of learning in general, as we accumulate information or skills, we naturally chunk related knowledge or skills together. This chunking strategy releases limited short-term memory capacity so that we can process information or function more efficiently. Dörnyei (2009) posits that this natural strategy is also evident in language learning. The initial language learning of toddlers is one example. Clustering of words can be seen in the functional communicative phrases of toddlers who use multi-word expressions that they copy from the language spoken around them before they actually understand the component words and structure (Myles et al., 1998; Peters, 1983). Such language use also suggests that word sequences can be stored and used holistically as if they are single words (Peters, 1983; Sinclair, 1991; Wray, 2002). Sequences that are argued to be stored holistically are identifiable by their phonological coherence when spoken, they may also appear more complex or longer than the surrounding language that the speaker uses. N. C. Ellis (2001, p. 39) explains chunks as the combinations of morphemes; for example, “the + ir” are two chunks which combine to a higher-level chunk: “their”. The more chunks that are combined in the higher-level chunk the less frequently it will occur, or the less useful it becomes. Therefore, we can expect that the lower level chunks are easier to retrieve because they are encountered more and are more useful. According to this concept, multi-word expressions are very high order chunks, and we can expect that a bigram (two-word sequence) will be easier to retrieve and use than a four-gram (four-word sequence). There is also evidence that if part of the chunk or multi-word expression is already known to the learner, they will be able to associate the additional component words more readily than if all component words were unknown (Kasahara, 2011; Y. Zhang, 2009). Also, congruency between multi-word expressions and their L1

equivalents can aid learning, while incongruency makes learning more difficult (Wolter & Gyllstad, 2011; Yamashita & Jiang, 2010).

Multi-word expressions have been described using various metaphorical labels. These “islands of reliability” (Dechert, 1983, p. 184) can be used as building blocks for the stream of speech. Sinclair (1991) suggests that most texts can be analysed using the Idiom Principle, that is to say that writers or speakers have chosen to use semi-preconstructed phrases for ease of communication when writing or speaking. Hasselgren (1994) coined the term *phrasal teddy bears*, to describe the familiar constructions that learners depend upon. N. C. Ellis (2012) describes prefab choices which aid language development but notes that such target-like choices tend to be lacking in L2 learner language. Prefab choices can also be imagined as “secure stepping stones, allowing speakers to plan the more creative stretches of speech” (Thomson et al., 2019, p. 406). It is possible that the chunks/multi-word expressions we use frequently are stored in the memory holistically (Tremblay et al., 2011). The next section introduces evidence to date for the effect that frequency has on chunking.

2.3.2 Frequency effects

In language learning, chunking can be explained through the effects of frequency, priming and automatization. The more that a language item is encountered, the more familiar it becomes, leading to productive confidence seen in faster retrieval and use. Lexical retrieval is known to put considerable demands on working memory, but as the learners’ familiarity with the words increases, an increase in speed of lexical retrieval (in this case retrieval of expressions) can be expected (Snellings et al., 2002). Bybee (2014) describes three effects of token frequency. The first is the *conserving effect* whereby repetition strengthens memory traces and makes the expression more accessible. The second effect is *autonomy*, which means the expression can stand alone without being constructed word by word. The third effect is the *reducing effect* whereby there is phonetical reduction often observed in common expressions.

It is possible that words prime each other and by virtue of their frequent use become more likely to be used together (Hoey, 2005). The language becomes easier to use with familiarity, and with enough practice perhaps proceduralised or automatized

(Gatbonton & Segalowitz, 2005). While the degree to which language items are automatized is very difficult to measure, the concept of knowing a multi-word expression so well that you can use it without conscious effort is certainly very attractive as a fluency development strategy, perhaps even necessary, because as DeKeyser (2001) states, “without automatization no amount of knowledge will ever translate into levels of skill required for real life use...” (p. 126).

2.3.3 Evidence of frequency effects from quasi-experimental studies

Before considering the evidence from various studies, it seems appropriate to clarify the difference between a true experimental study and a non-experiment/quasi-experimental study. A true experiment has randomly assigned participation, however most studies in applied linguistics use volunteers and class groupings for participants; making the studies quasi-experimental (Rogers & Révész, 2020). The majority of studies in the literature I will refer to can be categorised as quasi-experimental studies. Experimental and quasi-experimental studies include results from a control or comparison group which are used to infer what would have happened to the experimental group had they not received the experimental treatment. The definition of a control group can differ based on research domain (In’nami et al., 2020). Generally, in applied linguistics, a control group receives no treatment (only partaking in pre-and post-testing) (Loewen & Plonsky, 2016), whereas a comparison group receives an alternative treatment; for example the “best treatment currently available for the problem” (Shadish & Clark, 2004, p. 154). The current study introduced from Chapter Three is classed as a quasi-experimental study using a non-randomised control group; only involved in the pre- and post-testing but otherwise not exposed to the target structures. Now that the terminology has been clarified, I will describe evidence from quasi-experimental research regarding effects of frequency on learning.

Repetition or repeated encounters have long been associated with learning. In a meta-analysis of 26 L2 studies investigating the number of encounters for word learning, Uchiyara, Webb, & Yanagisawa (2019) found that massed learning conditions, where words are encountered repeatedly over a short span of time, had a greater correlation with word learning ($r = .38$) than spaced learning conditions ($r = .23$) where

encounters with target words are spread out. It is likely that repetition of the same text (through repeated reading, listening or role-playing) helps familiarize learners with the text content, which then enables learners to give more attention to unknown words or combinations. Knowledge of single words increases with repeated encounters (Horst et al., 1998; Webb, 2007). Collocation or multi-word expression knowledge also increases with repeated encounters (Webb et al., 2013).

Shadowing an audio recording of speech repeatedly can help multi-word expressions enter the phonological loop. Wood (2009) describes shadowing as “a technique often used in pronunciation teaching, in which a written text is read aloud while simultaneously listening to a recorded model” (p. 48). Shadowing was one of the activities reported by Wood (2009) which lead to fluency increases for a Japanese English learner in Canada, and the shadowing activity was repeated eight times for each model. Shadowing eight times might seem like a lot when time is limited. Shadowing four to five times is probably sufficient as Shiki, Mori, Kadota, & Yoshida (2010) found that shadowing between four to five times allowed low-intermediate to intermediate EFL learners in a university in Japan to increase their reproduction rate to a ceiling point. Frequency is important for making associations between words in one’s first language, however familiarity is more important for learning associations in one’s second or other language (Siyanova-Chanturia & Van Lanker Sidtis, 2019). Words that are frequently used together can be identified as chunks or a type of formulaic language. In this study, I am particularly interested in investigating and facilitating the acquisition of multi-word expressions by EFL learners.

2.4 Defining multi-word expressions

Formulaic language is estimated to account for up to 80% of spoken language (Altenberg, 1998), and approximately 50% of all discourse according to Erman & Warren (2000). In the past, formulaic language has been characterised as holistically stored units (Wray, 2002), however proving how a unit of language is stored in the brain is difficult which makes the use of such a criterion for identification of formulaic language somewhat restrictive. Recent consensus agrees that formulaic language “...need not be associated with holistic storage per se,” (Siyanova-Chanturia & Pellicer-

Sanchez, 2019, p. 3), though can be perceived “to have an identity or usefulness as a single lexical unit” (Wray, 2019, p. 267). A multi-word expression is a type of formulaic language. The term multi-word expression refers broadly to:

...(semi-) fixed, recurrent phrases, such as collocations (strong tea), binomials (black and white), multi-word verbs (put up with), idioms (spill the beans), proverbs (better late than never), speech formulae (What’s up), lexical bundles (in the middle of), and other types (Siyanova-Chanturia & Martinez, 2015, p. 549).

Multi-word expressions can be considered on a continuum from those that occur with high frequency in corpora to those that occur with less frequency. At the high frequency end of the continuum appear lexical bundles which occur at frequency ratios from 10-40 times per million words (Biber, 2009; Biber & Barbieri, 2007; Biber & Conrad, 1999; Byrd & Coxhead, 2010; Hyland, 2008). Highly frequent multi-word expressions such as these can be considered as general use and important for low proficiency learners to master because they occur so frequently, therefore their use can be expected to boost language fluency. At the other end of the frequency spectrum are idioms, which occur less frequently and often have specific uses. Use of idioms shows language sophistication, therefore they are suited for high proficiency learners to study. The likelihood of two words occurring side by side in a corpus is used to determine the strength of the word combination; this measurement is called the mutual information score. “The higher that score, the stronger the word partnership or collocation is” (Boers et al., 2014). Idioms are likely to have high mutual information scores as they feature lower frequency words and irregular grammar combinations, making their combinations more unique. On the other hand, higher frequency multi-word expressions such as lexical bundles combine high frequency words using regular grammar, making their word combinations inclined to have low mutual information scores. Since participants in my study had lower levels of proficiency, high frequency multi-word expressions such as *I think I will*, and *would you like to* would be the most useful type of formulaic language for them to acquire. First language speakers acquire multi-word expressions naturally through input from those around them and as mentioned earlier, toddlers are known to acquire expressions without any apparent recognition or analysis of the single

words that make up the multi-word expression (Peters, 1983). Later in their cognitive development, they will start to distinguish the separate elements of the expressions. We know less about how L2 learners acquire L2 multi-word expressions. However, adult L2 learners have a high level of analytic ability in comparison to a toddler and are likely to separate parts of a multi-word expression, rather than accept it as a single unit (Wray, 2004). Frequency of encounter with a language item is known to be important for acquisition, but for L2 learners the natural encounters that they have with L2 multi-word expressions are likely to be below the threshold for natural acquisition from input. Therefore, intervention is probably necessary to assist with the acquisition process, possible interventions are introduced in Section 2.5. Multi-word expressions are not only ubiquitous in language but have also been associated with fluency. The next section will review the literature showing the relationship between the use of multi-word expressions and fluency.

2.4.1 Associations between use of multi-word expressions and fluency

Familiarity with multi-word expressions contributes to receptive (listening and reading) language fluency, with many frequent word combinations able to be anticipated from trigger words or combinations (Conklin & Schmitt, 2012; Hoey, 2005; Pawley & Syder, 1983; Wray, 2002). Multi-word expressions take less time to read, an indication that it is easier for people to understand word sequences that are familiar and easily anticipated (N. C. Ellis et al., 2008; Siyanova-Chanturia et al., 2011; Underwood et al., 2004). The reduced processing load of multi-word expressions can also help to make the details which surround the familiar multi-word expressions easier to remember (see Tremblay, Derwing, Libben, & Westbury, 2011).

When it comes to producing language, the use of multi-word expressions assists the speed of production (in speaking and writing). For example, auctioneers and sports commentators who are known for their incredible speaking pace under the real time pressures of sales bidding or sports action often use multi-word expressions when speaking at speed, for example “any more bids”, and “last call” (Kuiper, 2004, p. 44). Perceptions of fluency are argued to depend on the ability to speak phrase by phrase rather than word by word (Wennerstorm, 2000). While highly frequent multi-word

expressions have been directly linked with reduced processing loads for native speakers, this fluency advantage is less likely to occur for non-native speakers. In fact, language learners are slow to develop knowledge and use of formulaic language despite the associated fluency benefits (Boers & Lindstromberg, 2012). Several intervention studies suggest that there is a link between L2 speakers using more multi-word expressions and speaking more fluently (Boers et al., 2006 (described in more detail in section 2.5.2); Wood, 2009 (described in section 2.6)). There is also evidence that use of multi-word expressions is positively correlated with speaking fluency and proficiency level in L2 speakers (Tavakoli & Uchihara, 2020).

Multi-word expressions are attractive for teaching because they provide bridges between lexicon and grammar (Bybee, 2008), especially useful for those starting out on their L2 learning journey. Knowledge and use of multi-word expressions is argued to provide known frames to experiment with in various situations and analyse as the learners' grammar develops (Myles, 2004). In a review of the evidence for the role of formulaic language in the L2 acquisition process, Wulff (2019, p. 30) writes "Formulaic sequences that are frequent and semantically transparent are likely candidates to serve as acquisition kick-starters." There are many formulaic sequences, idioms in particular, that contain irregular grammar. However, common regularly formed multi-word expressions offer potential exemplars for language learners as they strive to extend their grammar. In this section, the benefits to be gained from learning to use multi-word expressions have been laid out, which leads us to the next question, how can language learners learn to use multi-word expressions?

2.5 Ways to learn multi-word expressions

This section introduces some possible ways for language learners to learn and practice using multi-word expressions. Study abroad along with specific fluency instruction while abroad is considered first, followed by more classroom-based options such as noticing expressions in the input, learning lexical fillers as fluency strategies, retrieval and recall, memorisation, and rehearsal.

2.5.1 Study abroad

Study abroad is often linked with gains in language fluency. It has also been found that the longer second language learners live in an English immersion environment, the more their use of multi-word expressions increases (specifically two-word lexical bundles) (see Crossley & Salsbury, 2011). A study by Wood (2010) followed eleven study abroad participants who had either Japanese, Chinese (Mandarin) or Spanish as their L1 and were enrolled in an intensive English course at a university in Canada over six months. The participants were able to reduce their pausing time and extend their mean length of runs, and these significant fluency gains were associated with increased use of multi-word expressions. The course did not have any explicit focus on fluency development or formulaic language, so fluency gains and use of formulas were considered a natural product of the immersion environment. Participants were based in a homestay during the study or had experienced a homestay prior to the study. The homestay context was attributed a strong role in the development of fluency and formulaic competence in the participants (Wood, 2010). Not all language learners are able to experience an immersion language learning environment. Therefore, researchers have also investigated classroom interventions for learning to use multi-word expressions and develop fluency. Explicit instruction is sometimes necessary to notice the particularities of L2 forms. The next section will consider language teaching pedagogy and specific classroom interventions to help learners acquire multi-word expressions.

2.5.2 Noticing

One might assume that language learners will naturally encounter and learn frequent word sequences through normal classroom activities, without any need to explicitly draw attention to them. Indeed, incidental learning could occur for frequent word combinations if they are encountered and engaged with frequently enough, for example through extensive reading and other exposure. However, for those who are learning English as a foreign language, without an immersion experience, it is unlikely that there will be enough encounters for acquisition of multi-word expressions to occur without some explicit instruction.

It is argued that drawing attention to, or noticing multi-word expressions may be necessary for the learning process to start (cf. Schmidt, 2001). Learners are unlikely to notice the expressions without some kind of awareness raising intervention (Jiang, 2009). When a new language item is different in structure to the L1, as is often the case between English and Japanese, there may be need for explicit instruction, especially when there is no extra support from an immersion environment. If there are words in the expression that do not align to a meaning in the learners' first language, those words may be vulnerable to omission (left out of the speakers' utterances). In a study by Wray, (2004), a television presenter memorized a cooking recipe in a foreign language (Welsh) to present on a television program, the presenter was able to present the recipe very fluently in Welsh, despite not having previous knowledge of the language. However, in a delayed presentation of the same recipe, the presenter was found to omit small words from phrases presumably based on subconscious processes related to her L1. It seems therefore that teaching interventions for multi-word expressions should include explanation of the meaning or function of parts in order to help prevent omission of component words by learners.

The research for intervention studies experimenting with noticing of multi-word expressions has been carved out by Boers, Eyckmans, Kappel, Stengers, & Demecheleer (2006) who conducted a quasi-experimental study in an EFL context in Belgium, where they trained experimental participants to notice formulaic language (for example: "the vast majority", "make a difference") in their class materials (Boers et al., 2006, p. 260). This intervention was inspired by the lexical approach advocated by Lewis (1993) which premises that raising learner awareness of lexical phrases or chunks will prepare learners to independently notice and add such chunks to their linguistic resources. Over 22 class hours, the participants apparently developed strategic ability to identify and re-use multi-word expressions from L2 text prompts as evidenced in their final L2 re-tell task. As a result, the experimental group used more formulaic expressions in the re-tell task and were perceived as more fluent by blind judges than the comparison group. This study (Boers et al., 2006) is often cited as evidence for the association between L2 speakers using more multi-word expressions and being judged as more fluent as a result.

In an extension on the original study by the same research team using similar materials and procedures, Stengers, Boers, Housen, & Eyckmans (2010) provided an L1 text prompt for the L2 re-tell task to see whether experimental participants had committed more formulaic sequences to memory than the comparison group. An L2 text prompt was also given for comparison, to see whether they were able to re-use expressions from the prompt more strategically. In contrast to the first study, the researchers discovered that the experimental group did not use significantly more chunks than the comparison group in either re-tell task (L1 to L2, or L2 to L2). Participants' receptive chunk knowledge was also tested using a Discriminating Collocations Test (Eyckmans, 2009); however, no difference in knowledge gains were found between the experimental group and the comparison group, suggesting that the teacher-led noticing activities made little difference for these learners' development of general multi-word expression knowledge. The follow-up study suggests that the noticing intervention may not result in strategic ability to re-use text prompt multi-word expressions in speech as the first study suggested. The follow-up study did not report gains in fluency in relation to use of multi-word expressions so it neither supports nor contradicts the original Boers et al. (2006) result that linked greater use of multi-word expressions with greater fluency. The overall result suggests that simply noticing formulaic language from input (without rehearsal) is unlikely to result in acquisition (Stengers et al., 2010). Neither study reported a gain measurement for knowledge of multi-word expressions learnt through the interventions; such a measurement would give clearer evidence about the multi-word expression knowledge that was or was not acquired through the intervention.

2.5.3 Learning fluency strategies with lexical fillers

Teaching learners fluency strategies and lexical fillers such as 'let me think' over four weeks has been shown to increase fluency in a study abroad immersion context (Tavakoli, Campbell, & McCormack 2016). Participants with various L1s (predominantly Arabic, Kazakh and Chinese) were enrolled in an intensive English for academic purposes course in the UK (21 hours per week). Both experimental and comparison participants were trained in general speaking and listening skills, however, the experimental participants also received awareness raising instruction in fluency

strategies. Experimental participants listened to and evaluated a non-native speaker's picture re-tell recording for "speed, pausing and repair measures" and were taught to use lexical fillers such as "let me think," to avoid "unnecessary repair moves" (Tavakoli et al., 2016, p. 7). Experimental participants recorded their own picture re-tell speech, and after analysing it for fluency breakdown, recorded it again. While all participants improved on the fluency measures, the experimental group had a significantly greater speech rate gain (with a medium effect size) while the comparison group significantly improved their accuracy. The change in use of lexical fillers was not reported, so it is not possible to know whether the fluency increase was due to increased use of lexical fillers in particular.

2.5.4 Retrieval/recall

Encouraging retrieval/recall of target items from memory rather than simply reading the items or being given the answer has been shown to enhance learning and long-term retention (Roediger & Karpicke, 2006). Learning from this task type known as the *testing effect*, or elaboration rehearsal can occur in activities where students must recall or retrieve knowledge to complete a task. Retrieval can take different forms such as recognition or recall, receptive or productive. Recall is necessary if the learner needs to be able to productively spell the item (see Nakata, 2013, 2016). Productive recall ability develops from engagement or elaboration with the language item, stimulating a deeper level of processing, which leads to more reliable long term memory storage (Craik & Lockhart, 1972). An example of an activity that involves retrieval (recall) is asking learners to summarise a text from memory trying to re-use the words and expressions learnt (Jiang, 2009). Text reconstruction exercises, such as dictogloss (Wajnryb, 1990) where learners listen to a text and are tasked with taking notes and then reconstructing the text cooperatively, are another example of recall with elaboration. The dictogloss activity causes learners to focus on form as they produce written output through the reconstruction. Not only are the language items recalled but the user must thoughtfully consider how to fit the language into the surrounding text in order to achieve the intended meaning. In order to help learners notice multi-word expressions, Lindstromberg, Eyckmans, & Connabeer, (2016) adapted the dictogloss activity to highlight multi-word expressions in an academic article abstract for their Dutch

speaking university English major participants, who were enrolled in an English writing course. The experimental group were given the multi-word expressions in the order that they appeared in the audio text so they reconstructed the surrounding words, while the comparison group were simply given a blank piece of paper to reconstruct the full text on. Lindstromberg, Eyckmans, & Connebeer (2016) reported that this adaptation helped the experimental condition participants to recall and reproduce more text and more multi-word expressions than the comparison condition in an immediate post-test and a one-week delayed post-test. This adapted dictogloss appears to enhance learning of multi-word expressions and would therefore be a useful activity to use in the classroom.

A more recent study suggests the importance of involvement or elaboration on the meaning of the target expressions before using a dictogloss activity. Snoder & Reynolds (2019) compared two pre-dictogloss activities 1) a semantic elaboration activity, whereby participants wrote original sentences using the target expressions and, 2) a structural elaboration activity requiring participants to write phrases that rhymed with the target expressions. The semantic elaboration pre-task was found to promote greater learning based on receptive and productive post-testing. It seems that the more engaged learners can be with the meaning of the expressions before doing a dictogloss, the better chance they have of learning the expressions through the dictogloss activity. Learning has long been associated with memory; therefore, the next section looks at memorization as a strategy for learning to use multi-word expressions.

2.5.5 Memorization

Memorization is well known to assist the language learning process (Wray & Fitzpatrick, 2008), and simple phonological memorization can help a speaker speak fluently (Wray, 2004). Ding (2007) discovered that text recitation and imitation was an important language learning practice for three highly successful Chinese learners of English. Learners of Japanese in Singapore have also expressed appreciation for memorizing dialogues as a base and structure to use when faced with similar conversations (Walker & Utsumi, 2006). Memorised dialogues and learning of grammatical chunks in Japanese as a foreign language (at an elementary level in USA) has also been associated with increased language complexity, though not increased

fluency (Taguchi, 2007). In a study of Chinese learners of English, Yu (2009) demonstrated that committing a sequence to memory (difficult due to L1 transfer) resulted in greater procedural knowledge (being able to correctly translate from Chinese to English using the sequence). The comparison group that received explicit grammar instruction for the sequence however, achieved greater declarative knowledge (being able to correctly select the missing words from the sequence in a sentence context). The evidence reviewed here demonstrates the importance of memorization for building fluency and learning multi-word expressions. Since memorization is often achieved through rehearsal, the next section will review the research to date on rehearsal and its influence on fluency and use of multi-word expressions.

2.5.6 Rehearsal

Fluency, accuracy and complexity are all important for language performance but are also argued to compete for attentional resources (Skehan, 1996; Skehan & Foster, 1999). Bygate (2018), in the introduction to his edited collection “Learning language through task repetition” reflects that research to date on task repetition has shown that learners are likely to improve in some facet of their performance through iteration; however whether that be fluency, accuracy or complexity depends on the learner’s focus at the time. Activities that promote fluency are 1) communication focused, 2) use familiar language only, and 3) support and encourage performance at a higher than normal level (Nation & Newton, 2009). The 4/3/2 fluency activity where learners communicate the same information to different listeners in decreasing time limits (speaking for four minutes, then again but for three minutes with a new partner, and finally speaking for two minutes with a new partner) has been linked with increased fluency (Boers & Thai, 2017; Nation, 1989; Thai & Boers, 2016). The iterative nature of the 4/3/2 activity also provides opportunities for speakers to improve the quality of their spoken performance. For example, speakers can repair mistaken lexical choices and constructions in the second or third iteration. However, it is possible that opportunities to make such repairs or improvements based on previous performances may be lost if the increased time pressure precludes the opportunity to include them (Macalister, 2014). Thai & Boers (2016) demonstrated the difficulty of simultaneously increasing fluency, accuracy and complexity when they compared a 3/2/1 increasing

time pressure fluency activity with a 2/2/2 constant time condition. They found that while increasing time pressure encouraged fluency, keeping the time limits constant encouraged accuracy, complexity and fluency. In a replication of the 2016 study, Boers & Thai (2017) found that using a less familiar topic for the monologue inhibited participant speech. When the participants had less to say, the trend in results remained the same as in 2016 but the significant differences disappeared. They suggested building familiarity with a topic before using the repetition activity in order that learners might have ideas for what to say and get the most out of the activity (Boers & Thai, 2017). In another study of the 4/3/2 fluency activity, de Jong & Perfetti (2011) compared speaking on the same topic or speaking on different topics for the 4/3/2 fluency activity. They found those who spoke on the same topic were able to repeat more sentence structures (or multi-word expressions) within their monologue performances than those who had different topics, and such repetition resulted in a greater improvement in fluency (seen through phonation time ratio and pause times). If the 4/3/2 activity is being used in the classroom as part of a Four Strands approach, then as a fluency activity the content should be familiar. If the content is not familiar, then 4/3/2 is no longer being used in the intended way and cannot be classed as a fluency activity. Research to date has shown that rehearsal is extremely important for developing fluency or procedural knowledge. Indeed, acquiring a second language has been described as “a process whereby controlled, attention-demanding operations become automatic through practice” (McLaughlin, 1990, p. 125). In consideration of the evidence given above, it appears that foreign language learners stand to benefit from noticing, retrieving, recalling, memorizing and rehearsing highly frequent multi-word expressions.

2.5.7 The Four Strands balance

Most of the above studies focused on a particular intervention and its potential for multi-word expression learning and fluency development. This narrow focus is necessary in research endeavours that wish to measure the effect of a single variable and thus need to control for other variables. Of course, any of these singular activities such as noticing, or rehearsal used alone would make for a rather unbalanced and boring approach to language learning. In actual pedagogic practice, several interventions are

typically combined as part of a broader package of classroom activities, and it is therefore worth exploring the effectiveness of such a combination of activities.

Nation (2013) suggests balancing Four Strands of language learning activities. The Four Strands refer to 1) providing learners with ‘meaning focused input’ at an appropriate level for example with listening or reading texts, 2) ‘language focused learning’ where the meaning and form of new language items are given explicit attention, 3) ‘meaning focused output’ where learners try using the new language items in conversation, and 4) ‘fluency activities’ where learners practise using the new language forms in faster paced communication activities such as role-play and the 4/3/2 speaking activity. I am interested in developing speaking fluency through the use of multi-word expressions, and according to the Four Strands framework, I have to build up to the fluency activities by first providing learners with texts that they can comprehend (meaning focused input), followed by some explicit language focused learning where form, meaning and use are taught. Next, I need to provide opportunities to use the new forms in meaning focused output activities, and finally once learners are familiar with the word forms and uses, I introduce fluency activities where learners are re-using what has already been encountered. During fluency activities, I support and encourage learners to perform at faster speeds. It is important to be clear about the difference between fluency activities and speaking activities as Tavakoli & Hunter (2018) lament that fluency focused classroom activities are often confused by teachers as simply being ‘speaking’ activities. Fluency practice activities should 1) be communication focused, 2) only use familiar language, and 3) support and encourage performance at a higher than normal level (Nation & Newton, 2009). One example of an activity set focused on building fluency is ‘the daily fluency program’ (Millett, 2008). The daily fluency program is a quick 20-minute warm-up to a day of classes starting with five minutes ‘quickwrite’: writing on the topic of the day (a familiar topic); the focus is on quantity rather than accuracy. Next is ‘quickspeak’ where students speak to a partner on the same topic for two to three minutes without pausing, and then swap roles. The third activity is ‘quickread’ where students read as fast as they can and answer easy comprehension questions. Students record the time they take to read and their answer accuracy on a chart so that they can see their reading speed increasing.

While this approach promises to boost fluency, I would only have one 90-minute class once a week with my students, so the daily approach while attractive would not have been possible. There is also no focus on multi-word expressions in ‘the daily fluency program.’

2.5.8 Integrated skills (linked skills) programs

One way to achieve balance over the Four Strands is through an integrated-skills (or linked skills) program (Hirsh & Coxhead, 2009; Nation, 2014; Nation & Yamamoto, 2012), whereby learning focusses on the same content but it is re-used through different modes. For example, first, learners could listen to an audio text about robotic bees, and then they read the same text in order to answer comprehension questions. The activity just described is considered to be meaning-focused input (the first of the Four Strands balance). Learners might then be asked to notice new words and meanings that are glossed in the same passage, and they might have a word learning activity which would be language focused learning (the second of the Four Strands). If learners are then encouraged to reuse words from the text to write about their own ideas for designing robotic bees and present these ideas to a partner, this activity would be meaning-focused output (the third of the Four Strands). Finally, learners could be asked to present their ideas about robotic bee design to three other people, trying to improve the speed and accuracy of their explanation with every try, this activity would be fluency practice (the fourth of the Four Strands). The topic of robotic bees and related vocabulary and concepts are recycled throughout the stages of the lesson, this recycling of the topic links listening, reading, writing and speaking, therefore we call this a linked skills learning session. Linked skills lessons allow learners the opportunity to build knowledge and confidence with the topic content and related language which prepares them for fluency practice activities. Various learning activities can be used as part of a linked skills program; the examples given are just some of the possibilities. A linked skills course is argued to build fluency, but it does not focus on teaching or noticing multi-word expressions, so it would be interesting to compare such a language course with an explicit multi-word expression teaching intervention.

Gatbonton & Segalowitz (1988, 2005) recommend simulated communicative situations that create opportunities for repeated free use of appropriate multi-word expressions by learners (as opposed to repeating target multi-word expressions in a drill like fashion after the teacher). This repeated use is claimed to stimulate and support automatic retrieval and use of the utterances. Unfortunately, there are no results reported to support the claim that such activity sets would improve fluency and learning of multi-word expressions. That said, the evidence from other studies regarding repetition through various activities supports this claim in a broad sense. One example is a study reported by Wood (2009). The next section will describe the Wood (2009) fluency workshop study that combines learning activities that fit into the Four Strands balance and build up to fluency activities in a linked skills style. I chose to investigate the fluency workshop classroom intervention over other options due to the focus on speaking fluency, focus on multi-word expressions, similarity in the teaching program (once a week over six weeks) and claims of speaking fluency improvements from previous studies, despite issues associated with scale, design and context (McGuire & Larson-Hall, 2018; Onoda, 2014; Wood, 2009).

2.6 Explicit teaching and measurement of multi-word expressions

Wood (2009) reported an intervention study with a focus on learning multi-word expressions and developing speaking fluency, it was called the fluency workshop. This study is often cited as evidence that using more multi-word expressions increases speaking fluency, and for increase in use of multi-word expressions because of the intervention. In fact, this study has inspired a couple of adapted replications that suggest similar results, however, the studies are not without limitations and in order to understand the results of such an intervention, quasi-experimental research is needed. Wood (2009) reports a case-study, based on data from one Japanese participant enrolled in intensive study abroad intermediate level English language classes in a university context in Canada. The participant was also living in an English-speaking home-stay situation. The sequence of activities used in the intervention is shown in Table 2.1. The activities progressively built familiarity with multi-word expressions through repeated use in communicative activities making use of noticing, recall, memorization, and rehearsal over six weeks, with a 90-minute session once a week (nine hours in total). As

such, this fluency workshop fits nicely into the Four Strands balance set out by Nation (2007) mentioned earlier in the chapter and also conforms to a linked skills paradigm.

Table 2.1 Overview of activities in the fluency workshop

Stage (Wood, 2009, pp 48-50)	Four Strands (Nation, 2007)	Activity
Input	Meaning focused input	Listening for content
	Language focused learning	Listening and marking hesitations Linguistic and discourse functions of multi-word expressions
Automatization		Shadowing Mingle jigsaw
	Meaning focused output	Dictogloss Chat circle
Practice and production	Fluency practice	4/3/2 activity Record narrative and reflection
Free talk		Related topic free talk

For data collection, the participant's spoken narrative about past personal experiences was audio recorded pre- and post-intervention. The first narrative contained 530 syllables and lengthened to 760 syllables in the second measure six weeks later. Measures of fluency as seen in Table 2.2, increased from pre- to post-intervention, as did the use of multi-word expressions that had occurred in the teaching materials/model texts (for example: "almost every day, lots of, that's why") (Wood, 2009, p. 52).

Table 2.2 Results from fluency workshop

Fluency measure	% change from pre to post fluency workshop
Speech rate: syllables uttered per minute	13.8% (123.2 to 140.2)
Mean length of runs: total syllables uttered between pauses of 0.3 seconds or greater/# of runs	25.5% (5.1 to 6.4)
% syllables from formulaic sequences	10.6% (11.3% to 12.5%)
Mean length of formulaic sequences	40.7% (3.17 to 4.46)
Use of formulaic sequences from model texts	800% (2 to 18)

(adapted from Wood, 2009, pp. 50–51)

Formulaic sequences (multi-word expressions) were identified through phonological coherence, increased complexity, and syntactic or semantic irregularities. In the pre-intervention sample, multi-word expressions were mostly two to three-word collocations with straightforward functions such as temporal markers, for example “and then” and quantity markers, for example “most of them.” However, in the post-intervention sample, the multi-word expressions lengthened to include a wider range of functions and types such as cause and effect for example, “that’s why” and sentence builders for example, “one of my most vivid memories” (Wood, 2009, p. 52).

2.6.1 Limitations of the fluency workshop

As reported results were limited to one person, it is difficult to generalise or apply the results to a wider population. Without a broader sample or a control group, one cannot be certain that the fluency gains were a result of the fluency workshop alone. Perhaps simply participating in a general English class or studying abroad would have produced similar results. It could be that the immersion environment would have influenced fluency gains above and beyond the direct effect of the fluency workshop. Indeed, the strong influence of immersion was shown in Wood’s (2010) study mentioned earlier in the chapter. Despite the limited applicability of a case-study to a wider population, the Wood (2009) study has continued to be cited as evidence of the link that multi-word expressions have with fluency. There have also been attempts at extending and replicating this initial study to other populations and context.

Onoda (2014) ran a fluency workshop intervention based on the Wood (2001) model which was the basis for the 2009 model with more participants. Thirty second-year Japanese university students participated. Participants were majoring in English and also took other classes in English. The participants had an intermediate to upper-intermediate proficiency based on the Kanda English Proficiency Test, “which measures fluency and accuracy in university student oral production” (p. 130). As part of their Media English class, participants were guided to notice and use multi-word expressions over the course of one year from April through to January. Class activities at the input stage included watching a news story and a video of native speakers discussing the issue several times. Learners were encouraged to notice the formulaic expressions from the videos used for sharing opinions, checking understanding and summarising, for example, “I think,” “I agree,” and “that’s a good point” p. 130-131). For the automatization stage, class activities included shadowing and dictogloss. At the practice and production stage, class activities included the 4/3/2 fluency development activity and opinion sharing in pairs. A story-telling task was used to measure pre- and post-intervention fluency and accuracy. Participants watched a news story three times and could take notes while they watched, they were given several minutes to ask questions to the researchers about the content and then given another two minutes to prepare their notes and speech. Speeches were then recorded and analysed. Measuring trimmed speech rate, total unfilled pauses, frequency of pauses, and accuracy (number of errors in every 100 trimmed words), Onoda found an increase in both fluency and accuracy in the post-tests; indeed these were highly correlated “ $r = .89$ ” (p. 135). However, there was no control group, so it is impossible to say whether fluency improvements could be attributed to the fluency workshop intervention or simply a by-product of participants’ larger English study program. Nonetheless, Onoda’s study does suggest that fluency improvements from such an intervention can also be extended to an EFL context.

A quasi-experimental study by McGuire & Larson-Hall (2018) using similar activities to Wood (2009) compared results with a comparison group. They investigated whether fluency gains could be replicated in the ESL context of the United States of America over five weeks (eight classroom hours). The comparison group ($n = 8$) was taught using the same texts but were guided to learn vocabulary and grammar, while the

experimental group ($n = 11$) were guided to learn formulaic sequences. Both groups had conversational fluency speaking activities, with the experimental group being encouraged to use formulaic sequences in their conversations. Fluency was measured through a recorded dialogue between participants before and after the intervention. The experimental group showed greater gains in speech rate and mean length of runs and a medium effect for increase in use of formulaic sequences when compared with a comparison group. Overall, there was also a strong correlation between speech rate and use of phonologically coherent formulaic expressions. However, participant numbers were relatively low which throws doubt on the generalisability of the results. Also, knowledge and use of formulaic sequences that were taught to the experimental group were not reported separately, so it is not clear whether the intervention improved participants' knowledge and use of the formulaic sequences they had encountered and practised in class.

2.7 Need for replication

The above studies shine light onto potential classroom-based interventions for increasing awareness and familiarity with multi-word expressions which in turn might improve fluency. The evidence so often cited to support the link between fluency and multi-word expressions is not without limitations. Therefore, replication is necessary to ascertain the extent to which the results can be relied upon and generalized to other learners and contexts (Loewen & Plonsky, 2016). Replication is where we "...return to and repeat a previous study and compare what we discover with what was found or observed originally" (Porte, 2012, p. 2). Exact replication in the field of Second Language Acquisition (SLA) is extremely difficult to actualize due to differences in subjects and stimuli in most situations (Mackey, 2012). In this case, exact replication of Wood (2009) would not be terribly instructive because it would suffer from the same limitations that the original study had (no control/comparison group, lack of participants). On the other hand, an approximate replication would be able to address the limitations of the original study by introducing more participants and a control/comparison group to the investigation (as recommended in Thomson et al., 2019). An approximate replication is where changes are made to "nonmajor variables from the original study (in a way that allows for comparability between the original and

replication studies)” (Porte, 2012, p. 8). The goal of an approximate replication would be to confirm whether the fluency gains found in the original study could be “generalized to a new population” (Abbuhl, 2012, p. 301). I was interested to discover whether such gains could be seen in an EFL learning context in Japan where I teach.

Cultural differences exist between the ESL learning context in Canada and the EFL learning context in Japan. It is therefore possible that activities developed and used successfully by Canadian researchers and teachers with students studying within the Canadian cultural context may not be so successful in a different context due to different cultural norms and values. While cultural norms and values do not instantly change for a study abroad student, students are likely to attempt to fit into the new cultural context they find themselves in, while EFL learners remain in their cultural context so cultural preferences will likely play a stronger role in EFL contexts. For example, according to Hofstede’s cultural index comparisons, Canada has a low uncertainty avoidance index while Japan has a high uncertainty avoidance index (Hofstede et al., 2010, see pp. 192–194). The Hofstede finding describes Canadians as more comfortable with risk and Japanese as rather uncomfortable with risk. An example of how this permeates life in Japan is that many daily interactions are highly routinized which helps to reduce uncertainty. Parties where people mingle and talk to strangers may be commonplace in Canada, but in Japan people normally expect to have an introduction through a mutual friend before speaking with a stranger. Transfer this cultural preference to the classroom, and activities such as mingle jigsaw where learners freely wander the classroom and ask others to exchange expressions might work better in Canada than in Japan. Also Canada has a high index for individualism while Japan has a medium index (Hofstede et al., 2010, see pp. 95–96). We might expect, therefore, Canadians to be quite independent while Japanese will likely be concerned with maintaining group harmony. Transfer this to the group dictogloss activity, and learners in Japan might be more hesitant to share ideas in case they upset the group relationships in some way. Therefore, when transplanting classroom activities from the Canadian context into the Japanese context there is a need to check the appropriateness of activities and make adjustments as necessary. To this end, it would be useful to ask the end users how they feel about the replicated classroom activities in consideration of the

affective filter effect (Krashen, 1981). Listening to students' voices is vital to understanding learning outcomes (Coxhead, 2008). Various adaptations to activities would likely be needed in order to increase the appropriateness and effectiveness of classroom activities in the Japanese context.

In my teaching context, the class objectives of improving spoken English communication also made it necessary to adapt the fluency workshop activities from monologue to dialogue speaking activities, which are more useful for learners as conversation is the more common speaking format in day to day life. Using dialogue rather than monologue for speaking fluency assessment required decisions to be made about how to treat interactional features of language (such as shared pausing and overlapping speech) absent in a simple unidirectional speech flow (Tavakoli, 2016). The next section will introduce action research as a way to trial and adjust the fluency workshop classroom activities to fit into the Japanese cultural context and respond to learner feedback regarding the activities.

Chapter Three explains in more specific detail how I utilised this research method. The next section sets out the overall plan and research questions arising from the literature reviewed.

2.8 Summary of the literature

The literature argues that speaking fluency is an important and yet difficult to achieve aspect of L2 learning, especially for those learning in EFL contexts. The ubiquity of formulaic language and in particular multi-word expressions in corpus suggest that L2 learners need to know them. The associations between use of multi-word expressions and speaking fluency suggest that L2 learners would benefit from noticing and rehearsing multi-word expressions in interactive learning situations. The evidence supporting intervention is limited and therefore quasi-experiential research is needed to help verify the effectiveness of the fluency workshop activity set for learning to use multi-word expressions. In light of the literature to date, the next section will present the research questions for this research project.

2.9 Overall research questions

The fluency workshop concept holds latent potential for language learning situations where fluency is a goal. However, teachers are unlikely to adopt new teaching practices without strong evidence to show that they are more effective. Therefore, the overall research questions of interest were:

- 1) Does the fluency workshop intervention improve speaking fluency in an EFL context?
- 2) Does the intervention improve knowledge of multi-word expressions?
- 3) Does the intervention increase use of multi-word expressions in conversation?
- 4) What types of classroom activities are most appropriate for learning to use multi-word expressions from the perspective of EFL learners in Japan?

There were three studies, each with slightly different research questions which were intended to help answer the overall research questions. For ease of reference, Table 2.3 lists the research questions for each study.

Table 2.3 Overview of research questions for three studies

Study One Research Questions

-
- 1) Does repeated use of the target multi-word expressions in the experimental intervention activities increase meaning and form knowledge of target multi-word expressions?
 - 2) Does repeated use of the target multi-word expressions in the experimental intervention result in a greater increase in the use of multi-word expressions in speech than a linked skills program/control group?
 - 3) Does repeated use of the target multi-word expressions in the experimental intervention increase spoken fluency more so than a linked skills program/control group?
 - 4) Is there a relationship between fluency and use of multi-word expressions?
 - 5) Does focused instruction and practice with multi-word expressions reduce wider vocabulary learning?
 - 6) What adaptations are necessary for using the fluency workshop activities for a dialogue model and in an EFL context?

Study Two Research Questions

- 1) Is there an increase in multi-word expression knowledge in Study Two that supports the results from Study One?
- 2) Does multi-word expression use in Study Two increase more than Study One?
- 3) Does spoken fluency increase in Study Two more so than in Study One?
- 4) Does non-targeted vocabulary knowledge remain unchanged in Study Two as it did in Study One?
- 5) Which activities are considered the most useful by learners?

Study Three Research Questions

- 1) Is the increase in knowledge of target meaning and form found in previous studies replicable?
 - 2) Does practice of target multi-word expressions through different modes increase voluntary use in conversation?
 - 3) What does the intervention reveal about the relationship between use of multi-word expressions and fluency?
 - 4) Does greater vocabulary knowledge correlate positively with multi-word expression learning gain? Does focus on multi-word expressions affect wider vocabulary knowledge?
 - 5) Which activities are considered most useful by the learners?
-

3 Chapter Three: Research design and overall methodology

In this chapter, I introduce the research design and methodology for the study. The research was conducted using an action research framework with three iterative cycles of research, where each study made improvements based on the insights gained from the previous study. A quasi-experimental approach was taken comparing results between experiential and control groups. Any changes made to the base method introduced in this chapter are reported in Chapters Five (for Study Two) and Six (for Study Three). First, the differences between this study and Wood (2009) are described. Then the action research framework is introduced and considerations for ethics are described. The creation of experimental classroom materials and the selection process for the target multi-word expressions is described. Then the data collection instruments along with experimental and control group classroom activities are introduced. Next the procedures followed and finally the data analysis approach used for each data set is introduced. Also, please note the word *phrase* is used in this chapter interchangeably for multi-word expression, for example, the classroom activity *phrase instruction* refers to instruction regarding multi-word expressions. This terminology was used with the participants because I considered it easier for them to relate to and use than the longer multi-word expression term.

3.1 Differences between Wood (2009) and this study

In order to investigate the effectiveness of the fluency workshop it must be tested with more learners, and its outcomes compared with a control group (language learners not involved in the fluency workshop) (Thomson et al., 2019). Therefore, the current study takes the concept of the six-week (nine-hour) fluency workshop and puts it to the test in an EFL environment in Japan and compares results with a control group.

There are some differences between the fluency workshop that Wood reported in 2009 and this study. These are summarised in Table 3.1 and explained in detail below:

Table 3.1 Differences between Wood (2009) and the current study

Wood (2009)	Current study
Case-study	Quasi-experimental (incl. control)
ESL context in Canada	EFL context in Japan
Higher English proficiency	Lower English proficiency
Narrative model	Dialogue model
Narrative speaking tasks	Role-play speaking tasks
Phrase instruction in English only	Phrase instruction in English and L1
Shadowing frequency eight times	Shadowing frequency four times
Dictogloss	Dictogloss with MWEs printed
Decreasing time narrative 4-3-2	Decreasing time role-play 2-1.5-1
Spoken narrative recording	Spoken dialogue recording, MWE cloze test, LVLt test
No gauge of student sentiment to activities	Activity feedback surveys and focus groups

3.1.1 Case study vs. quasi-experimental study (including control group)

Whereas the original fluency workshop was a case study involving only one participant, this new fluency workshop was a quasi-experimental study conducted with a greater number of experimental participants and a group of control participants. Therefore, it was expected that the results from this study would be more generalisable as an indicator for the results likely to be seen in wider populations such as other EFL learners in Japan and potentially other EFL contexts.

3.1.2 ESL in Canada vs. EFL in Japan

The participants in the current study were living in a small town in Hokkaido, Japan and had much less exposure to the English language outside of class than the original participant in Wood (2009) would have, who was living in a homestay situation in Canada. Furthermore, because exposure to English is minimal outside the classroom in this EFL context, it is easier to attribute any observed changes in participants' task

performance to the teaching intervention rather than extraneous factors. Therefore, we can expect the results from this study to be able to be attributed solely to the intervention and to extend our understanding of the learning possible through the teaching intervention to EFL situations.

3.1.3. Higher proficiency vs. lower proficiency

The proficiency of the learner in the original fluency workshop was likely to be much greater than the new fluency workshop participants. This replication will therefore show how lower level proficiency learners respond to the fluency workshop.

3.1.4 Monologue vs. dialogue

The original workshop used monologues for models and recordings, while the new workshop used dialogues, since dialogue fitted well with the communicative goals of the course which participants were recruited from. Therefore, the narrative speaking activities of the original fluency workshop were changed to role-play speaking activities. This change in mode means that the results from this study will be relevant for conversation teaching scenarios, which are in high demand in many EFL contexts. More generally, of course, people engage in conversation more often than monologues, and so the present fluency workshop may align better with learners' real-life aspirations for L2 use.

3.1.5 Phrase instruction using English and L1

In the Wood (2009) study, phrase instruction would have been done in English only. However, in the current study, while English was the predominant language of instruction, Japanese was also used at times to enhance understanding since all learners were native or very proficient Japanese speakers (there were several international students from China participating). A shared language (L1) could be used because of the EFL context and shared language background. The L1 in such a context can be used as a resource to support L2 learning and understanding, while also making connections between known concepts/language and the new language (V. J. Cook, 1992; Storch & Wigglesworth, 2003; Tian & Macaro, 2012). All classroom instruction was given verbally, but also re-enforced visually in written form with use of PowerPoint slides.

3.1.6 Shadowing frequency eight times vs. four times

Shadowing was done eight times in a language laboratory in the Wood (2009) fluency workshop. However, in light of findings presented in the literature review (2.3.3) from Shiki, Mori, Kadota, & Yoshida (2010) that showed improvements stagnate after four repetitions, and the desire to not take more time than necessary in class with this activity, shadowing was done only four times in this study.

3.1.7 Traditional dictogloss vs. dictogloss with MWEs printed

The dictogloss activity in Wood (2009) was based on (Wajnryb, 1990) whereby students listen to an audio or oral reading of a text and take notes of the content, writing down key words or concepts on their blank note paper and then working in teams to reconstruct the text. However, in light of the positive findings regarding uptake of multi-word expressions in dictogloss by Lindstromberg et al. (2016), where the target expressions were pre-printed on the note paper. I decided to hand out paper with the expressions pre-printed on it and have the students listen to the audio and make notes for the surrounding text before sharing it with their partner and then another pair.

3.1.8 Decreasing time role-play 2/1.5/1

In Wood (2009), the 4/3/2 (in minutes) decreasing time speaking activity was used. Variation of the timing for this activity is not uncommon in order to adjust the activity to student capabilities, for example Thai & Boers (2016) used a 3/2/1 time variation. In this study, the times were shortened to 2/1.5/1, as these were the time segments that participants were able to fill with speech.

3.1.9 Data collection instruments

In the Wood (2009) study, the only data collection instruments were pre- and post-intervention spoken monologue recordings. I wanted to explicitly test uptake of the multi-word expressions, since speakers were not prompted to use them in the spoken recording, so a cloze test was added to this study as recommended in Thomson et al., (2019). Also, in Wood (2009), there was no independent measure of effect on other aspects of language learning such as general vocabulary, whereas I chose an external

vocabulary test to compare the effect of the experiment on learning of general vocabulary.

3.1.10 Feedback on activities from participants

This study also differs from Wood (2009) in that student sentiment was sought regarding the fluency workshop activities to discover what participants thought of the activities and also to get feedback for potential ways to improve the classroom activities or make them more contextually appropriate.

3.2 Cyclical action research approach

The current study method and materials differ from the original design of the Wood (2009) fluency workshop (described in Chapter Two) as has been outlined. I anticipated that I would need to make some adjustments to the classroom activities in light of the feedback from participants and my own observations.

In order to replicate and adapt classroom activities, space to trial adjustments was needed, along with a framework for reflection and identification of issues arising from contextual and cultural differences. Therefore, an iterative design utilising an action research framework was chosen. Action research is a research methodology where the researcher is also the practitioner or teacher participant in the classroom. The researcher identifies a problem in the classroom and then searches for possible interventions that might help to solve the problem (Burns, 2010). In many cases, action research is an iterative process in which an initial question or hypothesis is proposed, and an effort is made to address the question. Based on the results of the investigation, a new hypothesis might be put forward and investigated (Loewen & Plonsky, 2016, p. 1).

In action research, feedback regarding the classroom experience is collected from learners along with researcher observations. Such feedback on the classroom activities (the intervention) helps to critique the process of the intervention qualitatively rather than solely focusing on the quantitative pre- and post-intervention results (Loewen & Philp, 2012). Mandouit (2018) highlights the value of collecting feedback from students about how the learning activities and environment can be improved. Feedback can then be used to make changes to the classroom activities and instruction in an attempt to

improve them (Burns, 2010; Edwards & Burns, 2016). Though action research is not taken up in great numbers, the few English language teachers who engage with it are said to find it beneficial to their teaching practice (Borg, 2010). Action research is “context-specific, process-orientated and often described as cyclical” (Loewen & Philp, 2012, p. 63). One of the cited drawbacks of action research is that the results tend to be limited in application to the context in which they were found. I considered the action research method appropriate for checking the classroom activities for this intervention research. The action research framework was useful to gauge student responses to the classroom activities along with researcher observations in order to adapt the activities where appropriate, using iterative cycles to improve the classroom experience. Therefore, a cyclical design incorporating three cycles of action, reflection and adjustment were planned as illustrated in Figure 3.1.

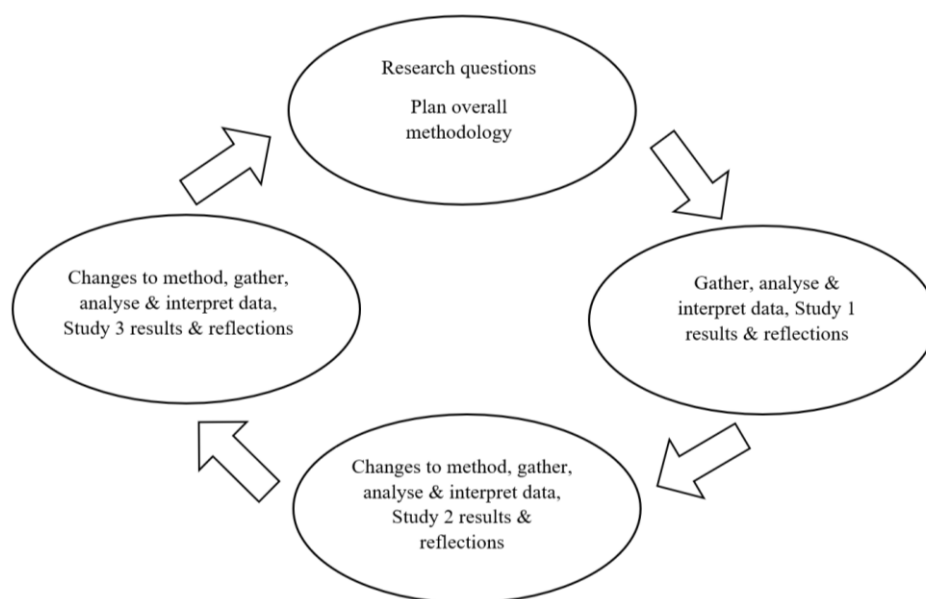


Figure 3.1 The action research cycle

Such a cyclical design fits well with the action research approach (see Burns, 2009), whereby a problem is identified (a lack of speaking fluency), an action plan or method implemented (the fluency workshop), results observed and analysed (through testing and surveys), and adjustments made in order to improve the teaching method. I

was both the researcher and the classroom teacher, a dual role typical of action research which “...involves taking a self-reflective, critical, and systematic approach to exploring your own teaching contexts” (Burns, 2010, p. 2). Therefore, I was interested in not only the research outcomes but also the classroom experience and how it could be improved for future use.

Using the action research framework, I set out to approximately replicate the Wood (2009) intervention in order to check whether the results would hold for an EFL context. My intention was that the results would tell us more about the validity of the Wood (2009) results and whether such results could be expected in an EFL context. The inclusion of a control group pushes this study into the quasi-experimental research quarter, allowing learning from the intervention to be isolated and measured. Therefore, this research project can be considered a fusion of replication, action research, interventional and quasi-experimental approaches. The next section will introduce the participants.

3.3 Experimental and control groups

Participants were recruited from intact EFL classes in an engineering university in Japan. The research project had three cohorts of undergraduate participants, a semester at a time, comprising both an experimental group and a control group for Study One and Study Three, while Study Two only had an experimental group due to lack of availability of a suitable control group. Study Two was therefore run to focus on changes to the classroom intervention based on Study One feedback, in preparation for Study Three where there would be more participants and a control group. Without a control group it is impossible to determine whether an improvement within the experimental group is due to the treatment or some other variable. Therefore, comparing experimental group results with control group results is important for determining or narrowing the possible causes for any changes after the experimental manipulation. Table 3.2 shows the participants over all three studies with their TOEIC score ranges. Each cohort (study) ran for six weeks, that is; six 90-minute sessions including pre- and post-testing, this time frame was replicating that seen in Wood (2009).

Table 3.2 Participants proficiency range over three studies

TOEIC score (IELTS equivalent)	S1 Exp (<i>n</i> = 52)	S1 Con (<i>n</i> = 35)	S2 Exp (<i>n</i> = 22)	S3 Exp (<i>n</i> = 65)	S3 Con (<i>n</i> = 51)
TOEIC 10 - 250 (IELTS 0 - 1.5)	13.5%	20.0%	31.8%	7.7%	7.8%
TOEIC 255 - 400 (IELTS 2 - 3.5)	57.7%	43.0%	36.4%	53.8%	58.8%
TOEIC 405 - 600 (IELTS 4 - 5)	25.0%	34.0%	31.8%	36.9%	23.5%
TOEIC 605 - 780 (IELTS 5.5 - 7)	3.8%	3.0%	0.0%	1.5%	9.8%

*S1 Exp = Study One Experimental Group, S1 Con = Study One Control Group

3.4 Ethics

I applied for ethics approval and it was approved at both the host institution in Japan; Muroran Institute of Technology, and also at Victoria University of Wellington. The ethics approval from Muroran Institute of Technology can be found in Appendix 1. The ethics approval from Victoria University of Wellington can be found in Appendix 2. An amendment to the ethics application approval with Victoria University of Wellington is provided in Appendix 3. This amendment related to offering financial incentives to participate in the focus group interviews. A further amendment is provided in Appendix 4, this amendment related to recruiting an assistant to timestamp the audio files of conversation between participants. An English translation of the information and consent page can be found in Appendix 5, the Japanese version which was used with participants is in Appendix 6. Volunteers for the focus group were given a separate information sheet and signed a separate consent form in Japanese allowing the discussion to be audio recorded, transcribed and analysed. The English translation of the focus group information sheet and consent form can be found in Appendix 7. The Japanese version is in Appendix 8. These information and consent forms were reused for Study Two and Study Three, the date by which a participant could withdraw from the study was however adjusted for each study. As both the researcher and the classroom teacher I made sure that students were aware that participation was voluntary, and that the research activities were not related in any way to their class grades. As the class teacher, I did my best to ensure that the activities we used in class would be at an appropriate level and in line with the class learning goals.

3.5 Experimental classroom materials

The participants for the experimental condition were recruited from an EFL oral communication course at the university in Japan. Therefore, the materials for the experimental condition were developed in line with the oral communication skill focus of the course; that is to build basic ability to communicate or speak using English. Three common travel related situations were selected as themes for the experimental intervention classes, since students would likely find use for them at some stage in their life, either travelling abroad or within Japan. The situations were: 1) deciding what to order from a menu at a *café* with a friend 2) asking, receiving and giving street *directions* and 3) making a walk-in *hotel* booking. A model dialogue for each situation was created by the researcher to use as a base for class materials; copies of the three model dialogues can be found in Appendix 9. The vocabulary used in the model dialogues was profiled to ensure use of high frequency vocabulary using Compleat Lexical Tutor v.8.3 (Cobb, 2020) with VP-Compleat (BNC/COCA 1-25K) lists. The first 2000 words in the BNC/COCA lists (Nation, 2012) provide good coverage of general service words for English learners (Dang & Webb, 2017). Table 3.3 shows the word frequency profile for each of the model texts; the predominant use of high frequency words (K1 and K2) means that the learners should not have had a problem understanding the vocabulary in the class materials.

Table 3.3 Word profile of model texts for Study One

Unit	Model no. of words	K1 token %	K2 token %	K3 token %	K4 token %	K5 token %	K6 token %
Café	105	95.2	97.1	98.1	100		
Directions	121	93.4	95.9	100			
Hotel	185	92.6	99.0	0		99.5	100

K1 = word appears within most frequent 1000-word list, K2 = word appears within most frequent 2000-word list and so on. The token % is the cumulative text coverage.

Further details on the vocabulary profiles can be seen with the model dialogues in Appendix 9. Conversational hesitation markers such as *um* and *mm*, and also people's names such as *Jon* were not included in the profile. Cognate words such as *omelette* that occurred in the café dialogue were counted as K1 words as they would likely have been familiar to the participants. An example of one of the model dialogues is given in Figure 3.2.

Figure 3.2 Example of model dialogue for café unit

P: Right, Jon. We have about an hour for lunch. The next meeting is at two thirty.

J: That's good. I usually only have time for a quick lunch, often at my desk!

P: Let's see what they have on the menu. What would you like Jon?

J: Mmm the salmon looks good, but so does the omelette. What are you going to get?

P: I think I'll have the chicken.

J: OK, I'll have the salmon.

P: Shall we get a salad to share?

J: Great idea, and are you going to get a drink?

P: Mm, I think I'll get a coffee, how about you?

J: Coffee, yeah me too.

P: OK, let's order!

I then audio recorded myself and another New Zealand speaker of English role-playing the model dialogues, so that the audio recordings could be used in the classroom activities. The model dialogues used in the experimental group classes each included 10 target multi-word expressions. Two weeks prior to the pre-tests, I used similar class activities with an airport themed unit to pilot the experimental activity formats with the students. This familiarisation with the activity formats was helpful for preparing the format of materials and instructions. Familiarisation with the activities also meant less

time would be required during the study for explaining new styles of activities to the learners. Unfortunately, I did not attempt to do any analysis of learner conversations at this stage, such an attempt may have helped foreshadow the difficulty I would eventually have identifying speakers within the conversations.

3.5.1 Target multi-word expressions (MWEs)

In this section I will explain how I chose the 30 multi-word expressions that were targeted in the classroom activities. There has been plenty of research that identifies the most frequent or useful phrases in English, resulting in academic purpose lists (for example; Simpson-Vlach & N. C. Ellis, 2010). There is, of course, a distinction between spoken and written mode, which affects which sequences or formulaic frames occur the most frequently in a corpus. Spoken formulaic sequences (combining function and content words) tend to be fixed, whereas written sequences tend to vary the content words used (Biber, 2009). In this study the focus is on spoken mode, so potential sources of multi-word expressions could have been Biber (2009) or Shin & Nation (2008), or the phrasal expressions list (Martinez & Schmitt, 2012). However, building up a communicative English curriculum from a list of phrases can be difficult due to lack of theme and cohesion; lists of phrases do not lend themselves easily to classroom integration (see Byrd & Coxhead, 2010). It is perhaps more practical for teachers who want to encourage learning of multi-word expressions to identify them from within classroom texts, or to integrate desirable multi-word expressions into their classroom resources. Therefore, I decided to use multi-word expressions that seemed useful and common that were occurring within my classroom materials and check them against corpora to ensure that they were in common usage.

In order to keep target multi-word expressions comparable, and to reduce overlapping, I decided to limit target expressions to no shorter and no longer than four words in length. Hyland (2008) justifies choosing four-word expressions for study as they are more common than five-word expressions and have a clearer range of structure and function than three-word expressions. Other studies have also set a length for formulas or expressions at four words (Biber & Conrad, 1999; Biber, 2009; Byrd & Coxhead, 2010). The expressions chosen contained a mix of function and content words

and are shown in Table 3.4. The multi-word expressions were used in pre- and post-tests and within experimental classroom activities (based on the fluency workshop model described in Wood, 2009).

Table 3.4 The thirty multi-word expressions targeted in the study

Café	Directions	Hotel
we have about an	How do I get	How many are in
I usually only have	Can you tell me	could you repeat that
have on the menu	When you get to	How do you spell
What would you like	turn left at the	How will you be
but so does the	until you get to	That will be fine
what are you going	You should see the	It is an extra
I think I will	right hand side of	Would you like to
I will have the	until I get to	let me think about
are you going to	is on the right	the details are in
to get a drink	have a good day	Thank you very much.

3.5.2 *Checking the usefulness of the multi-word expressions (MWEs)*

After multi-word expressions were selected from the classroom texts, the frequency of these expressions needed to be verified in order to establish that they were indeed used standardly in spoken English. The aim of the course was specifically aimed at building speaking fluency, so using a spoken corpus was important. The most appropriate corpus appeared to be the freely available spoken sub corpus (118 million words of unscripted conversation from American television and radio programs) of the Corpus of Contemporary American English (COCA) (Davies, 2013). It was initially considered desirable to have multi-word expressions that occurred over 20 times per million words in a spoken corpus since such frequent occurrence would be in line with previous definitions of lexical bundles; lexical bundles are identified through frequency and previous studies have used frequency ratios of 10 times per million words (Biber, 2009; Biber & Conrad, 1999), 20 times per million words (Byrd & Coxhead, 2010), and 40 times per million words (Biber & Barbieri, 2007; Hyland, 2008). Lexical bundles are the most frequent type of formulaic language, however due to their transparency and

often lack of clear L1 match, they are not salient and are therefore easily missed by learners. Additionally, lexical bundles are often not grammatically complete or idiomatic, but serve as building blocks for discourse (Biber et al., 2004). However, after checking the frequency data for the multi-word expressions in my class materials, it became apparent that many of the selected expressions did not occur at high enough frequency levels to be defined as lexical bundles in the spoken COCA. Perhaps the limited appearance of these sequences in COCA was a reflection of the very specific nature of the role-play contexts or the relatively small size of the spoken sub-corpus. The frequency threshold for selection was therefore set lower in order to allow for the context specific expressions to be included and to reduce the likelihood that all target expressions were already known by learners, thereby providing room for learning and avoiding any ceiling effects (Salkind, 2010) in the study. Some expressions that were highly frequent and therefore very likely to be known were also included in order to avoid zero scores in the pre-test (described later in this chapter), so that participants would not feel discouraged during or after testing. 29 of the selected expressions occurred between 0.03 to 124 times per million words, one further expression: *I usually only have* did not occur in the spoken corpus, but since it was considered useful for learning, it was searched for using the Google search engine, resulting in 1,200,000,000 hits, and this result was taken as evidence of common use. The inclusion of lower frequency sequences meant that I could not categorise this group of expressions as lexical bundles, hence the broader label of multi-word expressions.

The final 30 target multi-word expressions therefore included a mix of very frequent to less frequent expressions. The 30 expressions had predominantly question functions (11) and information giving functions (11). There were also instruction functions (4), response functions (2), a greeting, and a comparison function. A table of the 30 target multi-word expressions with frequency and function information can be found in Appendix 10.

The vocabulary that made up the 30 expressions was profiled using Compleat Lexical Tutor v.8.3 (Cobb, 2020) with BNC/COCA lists (Nation, 2012). The words that made up these expressions when profiled as single words are extremely common as shown in Table 3.5. All words that made up the expressions were from the most

frequent 1000 words in English except three words (*spell*, *repeat*, and *details*) these words were from the most frequent 2000 words in English; and one word (*menu*) was from the most frequent 3000 words in English. Shin & Nation (2008) argue that frequently occurring collocations made up of high frequency words should receive the same kind of attention as the high frequency single words, especially in speaking courses since collocations occur more frequently in spoken corpora than in written corpora. Therefore, this list of 30 multi-word expressions appears to be worthwhile for language learners to know.

Table 3.5 Frequency profile of words from the multi-word expressions

	K1 tokens	%	K2 tokens	%	K3 tokens	%
120 words	116	96.67%	3	2.5%	1	0.83%

3.6 Data collection instruments

This next section will describe the various data collection instruments used in this research to collect relevant information and measurements from the participants including surveys, pre- and post-tests, and instructions for dialogue recordings. An overview of the instruments used is shown in Table 3.6. The data instruments described here were used in Study One, Two and Three. Any adaptations made to the data collection instruments for Study Two and Three can be found in Chapters Five and Six.

Table 3.6 Overview of research instruments and their purposes

Measure	Purpose
Background survey	Compare participant language learning backgrounds & proficiency
Multi-word expression cloze test (pre & post)	Measure productive form and meaning knowledge of multi-word expression (declarative knowledge)
Dialogue recording (pre & post)	Measure ability to use multi-word expressions (procedural knowledge) and fluency (automatic knowledge)
Listening Vocabulary Levels Test (pre & post)	Measure general aural vocabulary knowledge
Activity feedback surveys	Compare perceptions of classroom activities (usefulness, enjoyment)
Focus groups/interviews	Gather further information about perceptions of classroom activities

3.6.1 Language learning background survey

A survey was conducted online in the classroom using Qualtrics to gather background information on participants in order to establish how homogenous the group was in terms of language learning background, and to assist in later analysis; triangulating and understanding any outlier performance in the pre- and post-tests. Multi-choice and yes / no selection questions were used to ask about extracurricular English study and study abroad experiences. First, participants read and responded to the information and consent form on Qualtrics. The language learning survey started on the page following directly after. There were nine basic questions with six follow-up questions appearing on the condition of an affirmative answer using skip-logic on the survey platform. An English translation of the language learning survey can be found in Appendix 11. The Japanese version which was used on Qualtrics can be found in Appendix 12.

3.6.2 General vocabulary knowledge

Proficiency is strongly associated with vocabulary knowledge (Stæhr, 2008). The starting proficiency of learners can affect their rate of uptake, so it needs to be measured. Based on chunking theory, we can hypothesise that learners with greater existing vocabulary will have an advantage for associating words into multi-word expressions (N. C. Ellis, 2001). Also, when testing an experimental and control group, tests are often focused on the phenomenon in question which tend to show the progress of the experimental group against the control group. It is not uncommon to see the experimental group improve on the target measure while the control group does not. However, it is possible that the control group improve at the same time on an untested measure. Measuring different aspects of language knowledge gives a broader and deeper perspective for analysis. It also seems wise to complement or triangulate internal tests (made specifically for the research project) with external tests, adding another yardstick with which to compare the analysis. To this end, it seemed helpful to survey participants' background language learning experience and proficiency and also administer an external test of general vocabulary knowledge. External because it is externally created and validated, and general vocabulary because vocabulary knowledge is related to language proficiency (Stæhr, 2008). Perhaps proficiency or general vocabulary knowledge would influence the effects of the intervention or perhaps the control groups would improve more on general vocabulary measures than the experimental group? I considered using the Vocabulary Levels Test (Schmitt et al., 2001), but decided to use the Listening Vocabulary Levels Test (LVLT) (McLean et al., 2015) because it was available in a bilingual format (English Japanese) and listening seemed the more appropriate mode over reading.

The LVLT test (levels 1-3 only, aligning with the most frequent 1000, 2000 and 3000 words in English) did not test the target multi-word expressions but rather general vocabulary at the highest three frequency levels. I thought therefore that the LVLT might be useful for potential triangulation with the internal measures, to see whether the control group might, for example, increase their aural vocabulary knowledge more than the experimental group, though materials for neither class focused on the words in the test. I hypothesised that the control group might not improve on the internal measures of

multi-word expression knowledge but that their vocabulary knowledge might increase as their course made use of a wider range of vocabulary. The LVL T test uses 24 multi-choice items per frequency level. The test words are taken from the first three lists of Nation's (2012a) BNC/COCA headword lists. Ability to recognize the words aurally and connect the sounds with the equivalent meaning in the Japanese language is tested. Participants listen to the audio and then choose the equivalent word in Japanese from the multi-choice options. The test was administered using Qualtrics.

For example:

Audio (this is only heard once; it is not printed) - Waited: I waited for the bus

- a) 食べた
- b) 待った
- c) 見た
- d) 寝た

Answer: b

3.6.3 Dialogue recordings (measuring of fluency)

A spoken role-play recording between participants was used to gauge spoken fluency and use of target multi-word expressions before and after the six-week treatment period. Use of multi-word expressions in the dialogue was completely unprompted and therefore as a measurement, the use of multi-word expressions was necessarily indirect. However, the production was natural and as such was informative about what had been learnt to the point of becoming procedural and or automatic knowledge able to support fluency. In the current study the dialogue role-plays were co-constructed between learners. Communication accommodation theory and verbal mimicry (Chartrand & van Baaren, 2009; Gasiorek, 2016; Giles, 2016) suggest that humans have an unconscious social response to converge with conversation patterns of the other which assists in building social solidarity. It could therefore be argued that the speaking proficiency of the interlocutor could influence the speech rate and other linguistic choices such as lexis to be lower or higher than the participant is capable of. Recent research on language testing has however shown no observable influence for

pairing speakers with different language proficiencies (Bennett, 2012; Davis, 2009). Therefore, participants were randomly matched together to perform the role-play according to the role-play instructions. The partners were kept the same for both the pre- and post-intervention recording. The instructions for recording the pre- and post-intervention dialogue in Study One can be seen in Appendix 13. The café menu which was used as realia for their interaction can be found in Appendix 14.

3.6.4 Multi-word expression cloze test

After the role-play recording, students were given a multi-word expressions cloze test to directly measure declarative productive meaning and form knowledge of the 30 target multi-word expressions (each four words long) using the quiz module on Moodle. Since declarative knowledge is known to precede procedural and automatic knowledge, we can hypothesise that learning of multi-word expressions that is revealed through this test, may not yet be able to be used in conversation. Therefore, this cloze test is likely to show more learning than analysis of spoken dialogue can. The text for the questions with answers can be seen in Appendix 15, together with a screen shot from the Moodle site to show how the test questions appeared to participants. The test format presented participants with a short one to two sentence context in English in which appeared the gapped multi-word expression. Based on previous experience using cloze tests with learners in Japan, I decided to provide the first letter for each component word in the target expression in order to help guide answers, and participants were to complete the missing parts. I had been told by previous learners that a cloze test that had no guiding first letter was simply too difficult to answer. Note, that single letter words such as ‘I’ and ‘a’ were given rather than gapped. A Japanese language translation of the expression was also given (recall that all participants were Japanese speakers). The Japanese translations were done by a professional Japanese translator and checked by me and another bilingual English Japanese speaker. The questions were automatically shuffled so each participant answered the questions in a different order. Below is an example of a cloze question.

For example: H__ d__ y__ s__ your name?

お名前の綴りをお伺いできますか？

Answer: How do you spell

3.6.5 Activity feedback surveys

Participant perceptions of usefulness and difficulty level of each classroom activity were considered likely to inform their preference for certain activities over others. Therefore, experimental participants were surveyed (using the Japanese language) in order to discover which activities were preferred and if there were any activities that should be altered or excluded in the second fluency workshop (Study Two). I translated the survey questions from English to Japanese and another bilingual English Japanese speaker familiar with Likert scale questions in Japanese checked the translations. The survey format is shown using a screenshot of the Qualtrics page in Appendix 16, an English translation is also provided. The survey format and wording were re-used for each activity, with only the name of the activity changed between surveys. The surveys were undertaken to help stimulate the action research reflection process so that I, as the researcher/teacher, could confirm or problematise the fit between the activities, the learning context and the learners.

3.6.6 Activity feedback focus group interviews

In order to further understand student sentiment regarding the activities, volunteers from the experimental group were recruited to participate in a 30-minute focus group interview outside of class time, to discuss their experience with the activities. Participants in the focus groups were shown the overall responses to the activity feedback surveys and asked about their experiences with the classroom activities as representatives of the experimental group. The focus group interviews were conducted using the Japanese language. The themes that came out of these discussions were noted for triangulation with other data to discover activities that were working well and others that were not.

This section has introduced the data collection instruments; the following section will describe the procedures for how the data collection instruments were used with the participants.

3.7 Procedures

This next section will describe the overall procedures used for the first of three iterative studies. These procedures cover the administration of surveys, pre- and post-tests, dialogue recordings and focus group interviews. The second and third studies also followed these procedures unless otherwise indicated in Chapters Five and Six. The first procedure was to inform participants about the study, ask for their consent to participate and administer the language learning background survey.

3.7.1 Informed consent and language learning background survey

Both the experimental and control classes were held in computer rooms, with each student sitting at a personal computer. Using computer classrooms allowed for all data collection to be done electronically and online via the Qualtrics survey platform and the Moodle Learner Management System. At the beginning of the six-week period of the first study, I gave a verbal explanation about the study in Japanese and English. Students then opened an online survey via the class Moodle page. The first page described the study in Japanese (the information sheet) and was followed by an online consent form (also in Japanese) with which students could freely agree to take part or not. These forms can be seen in Appendix 5 for English translation and Appendix 6 for the Japanese version. Those who elected not to participate still completed the tests and activities in class, but their data was not included in the analysis. Including the verbal explanation, this process took about 10 minutes in total. The third page on the Qualtrics survey instructed participants to wait for my signal before proceeding onto the next page. This signal was given once it was confirmed that everyone was at the same stage. The Listening Vocabulary Levels Test was next, and the procedure followed is described in the next section.

3.7.2 Listening Vocabulary Levels Test

The Listening Vocabulary Levels Test (McLean et al., 2015) was then introduced with a brief explanation and two practice questions. The test audio was played over the classroom speakers so that everyone could hear and do levels one through three of the test at the same time. Participants used computers to respond to the test items, selecting

the equivalent word in Japanese from multi-choice answer options of the English word which they had heard aurally. Each level took five minutes and after each level was completed there was a brief break where participants received automated feedback on their computer screen identifying their correct and incorrect answer choices. We then moved onto the next level as a class. In total the test took about 20 minutes (including transition times between the levels). The next section describes the procedure for the dialogue recordings.

3.7.3 Dialogue recordings (measuring of fluency)

After the listening test, participants were told to work with their partner (they were allocated seats based on randomly shuffled name cards at the beginning of the session, their partner was the person they were seated with). Instructions as seen in Appendix 13 were displayed on the screens at the front of the class and on each participant's Moodle page. Each pair was also given a laminated café menu as seen in Appendix 14. Participants were told to imagine that they were at a café with their friend and had an hour to eat. They were told to record their discussion as they role-played deciding what they would order using English in the café. Basic use of computer equipment for recording was demonstrated, and guidance given as needed. A forum within the class Moodle page was used to upload and post recordings. Audio was recorded either using Poodle; a Moodle plugin, or Audacity; where participants had to export their recording as an MP3 and upload it into the Moodle post. Participants were given four minutes to record and a timer was displayed on the board so that they were aware of the time available. I (as researcher/teacher) was available to help with any technical problems as required. After the role-play recording, all menus were collected. The next section describes the procedures for administering the multi-word expressions cloze test.

3.7.4 Multi-word expressions cloze test

After the dialogue recording, participants were instructed to open the multi-word expression cloze test activity on Moodle. Students only had 10 minutes to complete the test, so they were told to skip questions they did not know and move on to ones that they could easily answer. A countdown timer was displayed on the board so that

students were aware of how much time they had left. The time given to answer the questions was very limited, so it is likely that the test was tapping into easily accessible knowledge, that is, procedural knowledge. After the cloze-test, experimental and control class activities began. The next section describes the classroom procedures for the experimental group.

3.7.5 Classroom activities for the experimental condition

In the experimental intervention, the 30 multi-word expressions were taught across three thematic units (with 10 expressions being taught and practised per unit). Each unit took about 1.5 x 90-minute class sessions to complete, with classes being held once a week. Each class included around five of the workshop activities, and the following week the class would continue with the remaining five workshop activities for the unit. Experimental class activities are listed in sequential order in Table 3.7. These activities gave participants opportunities to read, hear, write and speak the target expressions. The sequence of the activities flowed from input through to practice and free production. I will now describe each activity as it was used in this study in detail. The activities are based on the Wood (2009) study but some differ as described earlier in section 3.1.

Table 3.7 Fluency workshop activities in sequential order

1	Listen with gist questions (10 mins)	6	Mingle jigsaw (10 mins)
2	Marking pauses (10 mins)	7	Role-play train (15 mins)
3	Phrase instruction (10 mins)	8	Decreasing time role-play (10 mins)
4	Shadowing (10 mins)	9	Record role-play (20 mins)
5	Dictogloss (15 mins)	10	Free related situation role-play (10 mins)

Approximate times taken for activity are given in parentheses.

3.7.5.1 Listen with gist questions

In this warm-up activity, the unit scenario (café/directions/hotel) was introduced with participants instructed to listen to the recorded model dialogue in order to answer several gist questions orally with partners, for example for the café conversation: What were the speakers talking about? What were they looking at? Participants were given the opportunity to think and then answer questions quietly with their partner to gain confidence in their understanding before I tried to elicit class level answers to the questions. The next activity was marking pauses.

3.7.5.2 Marking pauses

In this activity participants were given the transcript for the dialogue and instructed to mark any audible pauses while listening again to the audio. They were to indicate a pause by marking it with a diagonal line between words in the transcript. The audio was played twice, they then compared their pause markings with a partner, before seeing suggested answers on the screens at the front of the classroom. These answers were given as a guide and I told participants that variation to the answers was acceptable since the pauses were very difficult to distinguish. The main goal here was to show that the multi-word expressions did not have any internal pauses. For example, from the directions unit: So, /I turn left at the traffic lights /then go up Green Street /until I get to the Square (multi-word expressions are underlined here for the reader, however they were not underlined in the transcripts at this stage). This marking pauses activity was followed by phrase instruction.

3.7.5.3 Phrase instruction

In this activity, students were guided by the teacher to notice the target multi-word expressions by highlighting or underlining them in the transcript. Students were encouraged to take note of the meaning and function, which I explained verbally and presented on a PowerPoint slide. Students also copied my pronunciation twice for each multi-word expression, in order to familiarize them phonetically with the target multi-word expressions. Phrase instruction was followed by shadowing.

3.7.5.4 Shadowing

To further familiarize students with the flow, rhythm and pronunciation of the phrases within the context of the dialogue, students listened to the model dialogue again

using individual headsets and shadowed the dialogue at their own pace four times. I demonstrated shadowing the first lines, students were to speak the words just after the model audio with the transcript in hand for reference, providing visual reading assistance to the aural cues. The texts varied in length which affected in turn how long this activity took. The directions audio, for example was 50 seconds long, so shadowing four times took approximately five minutes. Shadowing was followed by the dictogloss activity.

3.7.5.5 Dictogloss

In order to take the pushed output from oral mode to writing mode, the well-known activity of dictogloss (Lindstromberg et al., 2016; Wajnryb, 1990) was used. Students received hand-outs with the ten expressions printed and space to fill in the surrounding text. They listened to the audio four to five times in total as a class; the first three times that they listened, they took notes and compared notes in pairs. After the fourth time listening, they worked to reconstruct the text with a small group of three to six people. Finally, they compared their reconstructions with the original transcript. An example of the café dictogloss worksheet is shown in Appendix 17. The dictogloss activity was followed by the mingle jigsaw activity.

3.7.5.6 Mingle jigsaw

The mingle-jigsaw activity (Wood, 2009) required students to hold the target phrases in their memory for a short time. Holding language in memory (phonological loop) for a short time and then producing it from memory without seeing the words (articulatory rehearsal), encourages learners to remember (retrieve) the language in chunks (making associative connections between the words) (N. C. Ellis, 2001). The mingle jigsaw was conducted as follows: I gave each person a multi-word expression printed on a small slip of paper which they were instructed to write down on their note paper and memorise. The small papers were then collected. I demonstrated how to ask another person what their phrase was and then to tell that person my phrase. After this interaction, we returned to our desks and wrote down the other person's phrase before repeating the sequence with a different person. After this demonstration, everyone stood up and mingled, telling others their phrase and hearing phrases from others. They returned to their desk after each interaction to write down the phrase while they could

still recall it. This process was repeated until the time finished or until they had collected and written down all 10 phrases in their list. The mingle jigsaw activity was followed by role-play.

3.7.5.7 Role-play train

The role-play pushed participants to practise using the phrases in the context of the model dialogue. Learners stood up and made two lines, facing each other as partners. They acted out the model role-play referring to their transcript as they role-played with their partner. After everyone had completed the role-play, one line of learners moved down the line so that everyone acquired a new partner and repeated the role-play. This sequence was repeated three times, and then the lines swapped roles, and the role-play was repeated another three times. In total, each person completed the role-play six times (three times in each role). I encouraged learners to depend less on reading from their notes in their third attempts by looking up from their notes and making eye contact with their partner when they were speaking. The role-play train was followed by a decreasing time role-play train which mirrored the role-play except that the time allowed decreased with each role-play and participants were not allowed to refer to their notes.

3.7.5.8 Decreasing time role-play

The decreasing time role-play (adapted from 4/3/2 activity, Maurice, 1983) encouraged learners to perform the role-play without notes and under time pressure. A large countdown timer was projected onto the two screens at the front of the class. Learners acted out the role-play in pairs without notes. The first time, they had two minutes; they then changed partners and completed the role-play a second time with only one and a half minutes; they changed partners again and acted out the role-play a third time, with only one minute to complete the role-play. They then switched roles and repeated the process. Therefore, there were six enactments per person (three times in each role). The decreasing time role-play was followed by an activity where learners recorded their role-play without notes.

3.7.5.9 Record role-play

Learners had to record their conversation without notes using a recording program on their computers called Audacity, and then export and save it before uploading it to a discussion post on Moodle. The process was always demonstrated and shown by the teacher, and assistance was given when necessary. The recording of the role-play was a fluency activity in that learners were again under time pressure and dialoguing the same scenario without notes. When the recording was completed, they were encouraged to listen back to it to check that their voice could be heard. In doing so, they had the opportunity to listen back to their performance and reflect on their fluency. The recording remained available online on the Moodle class page for students to listen back to whenever they liked. I did not however check whether they did so. The role-play recording was followed by a free role-play.

3.7.5.10 Free role-play

Finally, in order to give participants the opportunity to broaden their experience using the target phrases, they were given a related scenario to the model, for example, for the café unit, students could choose between deciding what to buy at a convenience store or at a clothing store. They were instructed to role-play freely with their partner, trying to use the target multi-word expressions in their conversation where possible. Students could refer to their notes if they wanted. After the free role-play the unit would be complete.

It should be noted that the experimental group were assigned weekly homework using the online video shadowing site English Central. Students were free to choose videos that were of interest to them, shadowing the audio and learning new words.

3.7.6 Classroom activities for the control condition

The control group experienced the pre- and post-tests, and they spent the same amount of time in English classes as the experimental group. However, they continued in their usual classroom activities and were not deliberately exposed to the target multi-word expressions for this study. The control group focused on two to three engineering themes, such as robotic bee design and purpose, over the same six weeks as the experimental condition. Topics were explored using various text genres such as email

and product explanations in a linked skills format. The control course unit sequence started with meaning focused input and language focused learning, in this case listening to a new audio text silently and filling in missing words or phrases on the class theme, reading the text and answering comprehension questions. Then, the control group completed meaning focused output activities (recycling the words and phrases from the model text into original writing) and fluency practice (sharing written work often verbally with others). In terms of output, the control group were asked to write emails and product design descriptions referring to a model text. They were encouraged to notice that they could change the content while re-using some helpful sentence structures. Some activities required learners to use their writing as a reference to then introduce the features of their design in speech to other learners. The control group were also given a new speed-reading text at the beginning of each class and were assigned extensive reading for homework (fluency practice).

Language learning through such linked skills progression has been recommended as an effective way to build fluency as students are able to progressively build familiarity with the content and related vocabulary (see Nation, 2014; Nation & Newton, 2008, p. 156). Therefore, while fluency was not the explicit aim of the control class, it was not neglected. The control group also had weekly assigned homework in the form of extensive reading. They were free to choose books to read that were of interest to them.

3.7.7 Activity feedback surveys (procedure)

In the final two class sessions for the experimental group, directly after the last experience with each activity while memory of the activity was still fresh, the activity feedback surveys were administered as described in 3.6.5. Each survey was administered online using the Qualtrics platform. Participants were presented with a statement “*Activity name* is too difficult,” and also “*Activity name* is useful for learning English”. Participants were to indicate the extent to which they agreed or disagreed with the statement on a five-point Likert scale. Participants could also write comments regarding the activities on the survey form. There were nine activities that feedback was gathered for in this manner.

3.7.8 Activity feedback focus group interviews (procedure)

Information was given to students at the end of class in weeks five and six of the study about the option to participate in a 30-minute focus group interview. The focus group interviews took place outside of class time, so a small monetary reward in line with the host institution's remuneration policies was offered to entice participation. Those interested in participating were asked to approach me after class to receive the information sheet and a form to complete in order to receive payment. Focus groups took place within two weeks of the study ending.

In Study One, four females; three Japanese from one class and one Chinese from another class volunteered to participate in a focus group. It was interesting that no males volunteered for the focus group, despite males making up the majority of participants. This female dominance could indicate a greater willingness by females to participate in discussion or could be because I (the teacher/researcher) was also female. Since these participants originated from classes with differing schedules; two focus groups were convened at separate times.

The focus group interviews were conducted as a discussion in Japanese because using Japanese put participants at ease to speak freely. Additionally, their English level was not sufficient to communicate the level of detail desired, so using Japanese was easier (see Hosoda, 2000 for use of L1 for interactional functions). I showed the focus group participants the response averages from the activity feedback survey results for each activity in turn and asked for reasons why students may have rated the activities as they did. I audio recorded the focus group interview with the participants' consent in order to analyse it at a later date. In the next section, I will explain how I organised and analysed the data collected through the various tests and measures.

3.8 Data analysis

In this section, I describe how I analysed the collected data. An alpha level of 0.5 was used to identify statistical significance. Distributions were checked for normality and are mentioned only when there was a significant deviation from a normal distribution. When there were ties in the data, the asymptotic significance p -value is reported and written as *asymp p*. Effect sizes are provided when applicable, these allow

comparison to be made across different data sets. I have used an online effect size calculator (Stangroom, 2018) to calculate Cohens d effect size. For Mann-Whitney (an alternative to the independent samples t -test) and Wilcoxon signed-ranks test (an alternative to the related samples t -test), I have used the effect size r , calculated as the standardized test statistic z divided by the square root of the number of observations (Pallant, 2007, pp. 224–225). The next section explains how the language learning background survey was analysed.

3.8.1 The language learning background survey

I downloaded responses to the language learning background survey into an SPSS file. I then removed those who had participated in both conditions; there were only a couple of students participating in both classes, this was due to the way that their personal class schedule had worked out. The English proficiency of the participants was self-reported through recent Test of English for International Communication (TOEIC) scores. TOEIC is a widely used English proficiency test in Japan produced by Educational Testing Service (ETS). For more information on the TOEIC test see <https://www.iibc-global.org/english/toEIC/test/lr/about.html>. The scores reported here are for the listening and reading test only. Analysis of the language learning background survey responses was not done in great depth unless there was an outlier or reason to look for more information to help understand the main pre- and post-test results. The next section explains how the Listening Vocabulary Levels Test was analysed.

3.8.2 Listening Vocabulary Levels Test (LVLT)

Responses to the LVLT were downloaded into a SPSS file and syntax used to automatically grade the test responses. Analysis was then done using SPSS to compare the groups' scores before and after the treatment. The next section explains how fluency was measured through the dialogue recordings.

3.8.3 Measuring fluency in dialogue recordings

There are several common measures of fluency that can be used to analyse speech samples. I chose to investigate speech rate and mean length of runs in Study One.

Phonation time ratio and articulation rate were also investigated in Study Two and Three. These measures of fluency are defined and explained below.

Speech rate (syllables uttered per minute) is a global composite measure which indicates speed of delivery (inclusive of pausing) (see Tavakoli et al., 2016). Pausing or hesitation is not necessarily an indication of dysfluency as it is common before less frequent or predictable words and can be a signal to the listener to listen carefully. Research on British English listeners has shown that listeners are more likely to recognise words preceded by hesitation, than words spoken without any hesitation (Corley et al., 2007). Whether or not foreign language learners of English pay more attention to words preceded by hesitation is beyond the scope of this study, however when we measure fluency and include pausing in the measurement it is worth noting that hesitation is not necessarily problematic.

Mean length of runs is another measure of fluency, it is calculated as the average number of syllables uttered between pauses. Generally, the longer the mean length of runs, the more fluent the speaker is considered to be. Tavakoli (2016) found that the mean length of runs tended to be longer in dialogues than in monologue.

Phonation time ratio is another measure of fluency which is the “percentage of time spent speaking as a percentage proportion of the time taken to produce the speech sample” (Towell et al., 1996, p. 91). Evidence of proceduralisation is found in the combination of an increase in mean length of runs with a stable phonation time ratio. Meaning that longer speech sections are occurring without an increase in pausing time (Towell et al., 1996). Articulation rate measures the pace of actual speaking when pauses are excluded. It is calculated as syllables divided by speaking time (excluding pauses) x 60 (Towell et al., 1996).

The literature often describes and measures fluency in the context of monologic speech. Dialogic speech can be measured in a similar way however conversational features such as overlapping speaking and shared pausing also need to be considered. For example, to which speaker should pauses that occur between speaking turns of different speakers be associated with; the speaker who has just finished speaking or the speaker who is about to speak? Tavakoli (2016) split such pauses between the two

speakers so that each speaker was attributed half of that pause time in their fluency measurement, I decided to follow this precedent from Study Two.

Another decision that needs to be made is after how many fragments of a second is non-speaking time counted as a pause. Pauses have been defined as non-speaking time longer than .2 seconds (N. de Jong & Perfetti, 2011), .25 seconds (Tavakoli, 2016), .3 seconds (Wood, 2009) and .4 seconds/400 milliseconds (Freed, Segalowitz, & Dewey, 2004). I decided to count non-speaking time longer than .25 seconds as a pause from Study Two, following the more recent precedent.

Filled pauses also need to be considered, should they be measured as speech or pause? I decided to include filled pauses (for example *um*, *ah*, *er*) and repetitions as speaking time. Repetition of words could be considered as a dysfluency indicator, but it can also be a technique for gaining planning time. Acquiring planning time frees up cognitive resources to improve performance (Bygate & Samuda, 2005; R. Ellis, 2005). Therefore, one has to decide whether to focus on the full speech sample (untrimmed), or a trimmed sample which removes repetitions and filled pauses. I decided to analyse both trimmed and untrimmed samples.

The spoken data was sorted so that recordings were only subject to analysis if they met the following criteria: 1) both speakers had given consent to participate in the study, 2) the pair of speakers were the same for both pre- and post-test recording, and 3) the audio recording was of sufficient sound quality to analyse. The recordings from those who had different partners were not included in the analysis in order to reduce noise in the data, since partner dynamics could affect speech performance. As it was difficult to be sure which individual was speaking in Study One, the transcriptions were not split by individual but rather transcribed as dyad interactions and analysed as a whole conversation. Therefore, each case number in the statistical analysis for Study One actually combines the utterances of two people.

3.8.3.1 Transcription

First, I transcribed the spoken dialogues verbatim into notepad++ software. All numbers had to be spelt out in the transcripts along with words such as *okay* rather than the abbreviated *OK* in order for syllables to be counted correctly by the software.

Following the initial transcription, a copy of the file was made and all repetitions, fillers (such as um, arh etc.) and false starts were removed to create trimmed transcriptions. However, repeating or echoing back what a speaker had just heard the other say was not removed as this type of repetition could be interpreted as confirming meaning or some other pragmatic communicative act, differentiating it from straight repetition. Pauses were marked manually in the transcript by a new line, and every speaker change was also shown by a new line. The lines in the file could then be automatically counted and this number became the number of runs.

3.8.3.2 Analysing the transcripts

Word count, syllable count and number of runs were calculated by copying and pasting the speech transcription into the online site <http://www.syllablecount.com/#results>. The mean length of runs was then calculated by dividing the total number of syllables by the total number of runs. Speaking time was counted until after the last utterance and measured in seconds using Windows Media Player in Study One (PRAAT software (Boersma & Weenik, 2017) was able to be used for Study Two and Three because files were saved in WAV format rather than MP3). Speech rate was calculated by dividing the total syllable count by the seconds taken and then multiplied by 60 to give a syllables per minute speech rate. Data was entered into Microsoft Excel and then copied across to SPSS for further analysis.

3.8.3.3 How many multi-word expressions should we expect?

The role-play task did not force the use of target multi-word expressions nor did it prompt their use in any way, so it was entirely possible to work through the role-play using alternative expressions or language. Therefore, measuring use of multi-word expressions in the recorded dialogue was an indirect test of the ability to use the target multi-word expressions. However, the target multi-word expressions were considered common enough to be natural language choices in the role-play task. It should be noted though that since the pre- and post-intervention dialogues were based on a café scenario, it was unlikely that participants would use multi-word expressions specifically related to giving directions or booking a hotel, which experimental participants had also been exposed to. Therefore, the pre- and post-intervention dialogues were only likely to draw out 10 of the 30 possible target expressions.

3.8.3.4 Counting use of multi-word expressions

In order to count the use of multi-word expressions, I used the Antconc (L. Anthony, 2014) program. Using Antconc, I could scan the transcripts for multi-word expression use, that is, sequential use of the target multi-word expression words. I used the concordance word search function, first searching for use of the full four words, then by three-word and two-word combinations. Use of the full four words of each multi-word expression was rare, so it was important to measure partial use to gather information. Each target multi-word expression had four words, so one point was assigned for the use of each word used in combination of two or more words to express the target multi-word expression meaning or a related meaning, enabling partial use of multi-word expressions to be measured (Barcroft & Rott, 2010 also used partial counts of words to assess learning). Therefore, if someone were to use each of the 10 target multi-word expressions completely in the dialogue, they would score 40 points for multi-word expression use. The Antconc search not only revealed target like use of the multi-word expressions but also related use of the multi-word expressions. For example, if the target multi-word expression was, *I think I will*, and a participant said *I think that one*, then the utterance would be counted as two points since it consisted of two of the four target words in sequence (*I think*). If words from the multi-word expressions overlapped, they were only counted once. For example, *what would you like* and *would you like to* were sometimes combined into a five-word expression, so in such an instance the five words were only counted once as a five-word expression. *Do you* (from *how do you spell*) was often used, but not counted as *do you* if used in combination with *do you like*, in this case it was counted as *you like*, to avoid double counting (from *what would you like*). If a multi-word expression was split over two runs, it was still counted. That is to say that prosodic coherence was not a criterion for counting use of multi-word expressions. The next section explains how the cloze-test results were analysed.

3.8.4 Multi-word expressions cloze test (data analysis)

The multi-word expressions cloze test was automatically scored by the quiz program on Moodle. Only the exact target words spelt correctly and in the correct place order were counted. A full correct four-word expression was worth 1 point. If only one

word was correct out of the four, then the score would be .25 for that multi-word expression. For expressions with a single letter word, the score was split over the other three words. For example, the target expression ‘how do I get’, was scored as .33 for each correct word as ‘I’ was given. Therefore, if a response were ‘how do I go’, then the score would be .67 as ‘go’ was not a target word, and ‘I’ was not counted. The first letter was given, so only the following letters were necessary, but if the whole word was entered including the first letter, this was also marked correct, since entering the first letter again was an easy mistake to make, even for highly proficient English speakers who the test was tested on at the design stage. Therefore, the total score for the test was out of 30 points. Responses were downloaded into Excel and then transferred into SPSS for analysis.

3.8.5 Activity feedback surveys (data analysis)

Responses to the activity feedback surveys were downloaded and collated into SPSS and Excel for analysis. The results were collated into three levels of agreement for easy explanation. That is, strongly agree and agree are simply reported as agree, and strongly disagree and disagree are simply reported as disagree. The middle choice is reported as neutral. The responses are reported as percentages of those who agreed or disagreed with the statement. Comments from the survey were also collected and have been used to help explain the survey results or add further information to the comments from the focus group where helpful.

3.8.6 Activity feedback focus group interviews (data analysis)

I listened to the focus group audio recordings and noted the main points and topics in English, using ELAN software. Quotes of interest relating to the classroom activities were transcribed. The notes and quotes from the focus group interviews were used to add additional detail and depth to help shed light on the activity feedback survey responses. Quotes from the focus group participants that added depth to the learner perspective were translated into English and were included in reporting to support and further explain survey results.

3.9 Summary of overall methodology

In this chapter I have laid out the differences between this approximate replication and the original Wood (2009) study. I have explained the procedural framework of the study using the action research paradigm. The ethical considerations have been explained particularly considering my dual role as researcher and classroom teacher. My method for selecting multi-word expressions for classroom focus has been explained. I have introduced the data collection instruments (surveys and tests) that I used in these studies. The class materials used in the studies have been introduced, explained, and justified. I have also explained the procedures for classroom activities and testing, and how I conducted the data analysis of test and survey results. The results from Study One are presented in the following chapter; Chapter Four.

4 Chapter Four: Study One

This chapter reports the results from Study One, which was carried out using the methods and tests described in Chapter Three. An article has been published based on the results from Study One (Thomson, 2018), but the reporting here is more detailed and does not copy the wording from the article. Results will be reported in relation to each research question below, first by comparing the experimental and control group, and then reporting on any changes within the groups. Where applicable, the results from this study are then also compared with the results from Wood (2009).

4.1 Research questions

In this section the six research questions that I set out to investigate through Study One are listed.

- 1) Does repeated use of the target multi-word expressions in the experimental intervention activities increase meaning and form knowledge of target multi-word expressions?
- 2) Does repeated use of the target multi-word expressions in the experimental intervention result in a greater increase in the use of multi-word expressions in speech than a linked skills program/control group?
- 3) Does repeated use of the target multi-word expressions in the experimental intervention increase spoken fluency more so than a linked skills program/control group?
- 4) Is there a relationship between fluency and use of multi-word expressions?
- 5) Does focused instruction and practice with multi-word expressions reduce wider vocabulary learning?
- 6) What adaptations are necessary for using the fluency workshop activities for a dialogue model and in an EFL context?

4.1.1 Participants

Only data from those who consented to participate are included in the analysis and reporting. The language learning background survey was administered at the very beginning of the fluency workshop for both the experimental group and the control

group. The participant population was made up predominantly of Japanese males aged 18-19 years old. The first language of most participants was Japanese; however, one Chinese student was included in the experimental group, and another Chinese student was included in the control group. Both Chinese students had good Japanese proficiency and were studying all their undergraduate courses through Japanese language, so it can be assumed that they understood all Japanese language used in the study. Participants were prompted to self-report their most recent TOEIC score range. English proficiency spread as identified by TOEIC score is shown in Table 4.1, with the majority scoring between 255 and 600. A score range of 255-400 can be described as elementary proficiency, while 405-600 can be described as elementary proficiency plus. The score range goes from 10-990 (for more details see <https://www.examenglish.com/TOEIC/index.php>). An independent *t*-test showed that there was no significant difference between the average TOEIC score ranges reported by the experimental group ($M = 2.19$, $SD = .72$) and the control group ($M = 2.20$, $SD = .80$); $t(85) = .047$, $p = .96$, $d = .013$. Therefore, the groups appeared to be roughly equal in terms of English proficiency.

Table 4.1 Study One participant numbers by group with TOEIC score range

TOEIC score (IELTS equivalent)	Experimental ($n = 52$)	Control ($n = 35$)	Total ($N = 87$)
10-250 (IELTS 0-1.5)	14%	20%	16%
255-400 (IELTS 2-3.5)	58%	43%	52%
405-600 (IELTS 4-5)	25%	34%	29%
605-780 (IELTS 5.5-7)	4%	3%	3%

Prior to analysis of results for each variable, normality of distribution was checked by looking at normality plots of the data and normality tests, such as Levene's test, and tests for kurtosis and skew. The following results will only mention normality testing if the normality of the data distribution was in doubt. Statistical significance recognition was set at an alpha level of .05.

4.2 Knowledge of target multi-word expressions

The target multi-word expression cloze test, a direct test of productive knowledge for the 30 target multi-word expressions, was administered at the beginning of the six weeks for both the experimental group and the control group. For more information on the test see 3.8.4. The highest possible score on this test was 30. The average pre-test score for the experimental group ($n = 44$) was 8.95, $SD = 3.42$, and the average for the control group ($n = 29$) was 10.0, $SD = 3.89$. An independent t -test showed no significant difference between the groups $t(71) = 1.21$, $p = .229$, $d = .29$, 95% CI [-.68, 2.77]. At the end of the six-week period, the same test was administered again. An independent t -test showed a significant difference and large effect between the experimental group ($n = 44$) average score of 16.6, $SD = 5.19$, and the control group ($n = 29$) average score of 12.6, $SD = 5.07$; $t(71) = 3.28$, $p = .002$, $d = .79$, 95% CI [-6.49, -1.58]. The next section analyses gain scores for the cloze test.

4.2.1 Multi-word expression cloze test gain scores

Because the two groups' pre-test scores were slightly (albeit non-significantly) different, the difference between pre- and post-multi-word expression cloze test scores was calculated as a gain score, for a further comparison of the two groups' learning rates. An independent samples t -test showed that productive knowledge of target expressions increased significantly more for the experimental participants ($n = 44$) $M = 7.63$, $SD = 4.12$, than the control participants ($n = 29$) $M = 2.55$, $SD = 3.01$, $t(71) = -5.71$, $p < .001$, $d = 1.41$, 95% CI [-6.86, -3.31]. It is clear from this result that the experimental group outperformed the control group in the post-test, despite starting from an equal knowledge position in the pre-test. The next section analyses within group differences for the cloze test from pre- to post-intervention.

4.2.2 Multi-word expression cloze test within group comparisons

A paired samples t -test showed that the experimental group ($n = 44$) improved their score significantly between their pre-test ($M = 8.95$, $SD = 3.42$) and post-test ($M = 16.6$, $SD = 5.19$), $t(43) = -12.3$, $p < .001$, with a large effect size $d = 1.74$, 95% CI [-8.9, -6.4]. The control group ($n = 29$) also improved significantly from pre- ($M = 10.0$,

$SD = 3.89$) to post-test ($M = 12.6$, $SD = 5.07$), $t(28) = -4.56$, $p = < .001$, with a medium effect size $d = .564$, 95% CI [-3.69, -1.40].

4.2.3 Answering RQ. 1

Therefore, in response to research question one, the results of the direct test of multi-word expression knowledge show an advantage in multi-word expression learning for the experimental participants, suggesting that the fluency workshop is able to increase productive knowledge of multi-word expression form and meaning. This result will be discussed in further detail in Chapter Seven, the discussion chapter.

4.3 Multi-word expression use

This section reports on the use of multi-word expressions in the pre- and post-intervention dialogue recordings. A list of the multi-word expressions used (spoken) by participants in the dialogue recordings in Study One can be found in Appendix 18. The use of multi-word expressions in the role-play recordings was analysed as described in Chapter Three with each word from an expression scoring a point. The average use of multi-word expressions in the pre- and post-measure dialogues is shown for each group in Table 4.2. Notice the difference between the groups widens in the post-test scores.

Table 4.2 Multi-word expression use in the pre- and post-measure

Multi-word expression use (by word count)	Pre-test <i>M</i>	<i>SD</i>	95% <i>CI</i>	Post-test <i>M</i>	<i>SD</i>	95% <i>CI</i>
Experimental (<i>n</i> = 15)	6.20	4.07	3.53, 8.87	11.0 +77.4%	7.34	7.55, 14.5
Control (<i>n</i> = 8)	5.00	6.41	1.34, 8.66	6.88 +37.6%	3.98	2.16, 11.6

The use of multi-word expressions appeared to follow a normal distribution in the pre-measure dialogue for the experimental group and the post-measure dialogue for both groups, although normality was doubtful for the control pre-test dialogue where there was an outlier and a significant result on Kolmogorov-Smirnov and Shapiro-Wilk

tests of normality. Using an independent samples *t*-test on multi-word expression use in the pre-measure dialogue, no significant difference was found between the experimental group ($n = 15$) ($M = 6.20$, $Mdn = 12.0$, $SD = 4.07$) and the control group ($n = 8$) ($M = 5.00$, $Mdn = 8$, $SD = 6.41$); $t(21) = .551$, $p = .588$, $d = .224$. Due to concerns over normality in the control group, this analysis was supplemented with the non-parametric Mann-Whitney U ($U = 78.5$, exact $p = .238$, $z = 1.22$, $r = .253$), which also recommended retaining the null hypothesis. Therefore, it can be assumed that the two groups were using a similar quantity of multi-word expressions in their dialogues prior to the experiment.

An independent samples *t*-test was then used to compare multi-word expression use between the groups in the post-test dialogues; and no significant difference was found between the experimental group ($M = 11.0$, $SD = 7.34$) and the control group ($M = 6.87$, $SD = 3.98$); $t(21) = 1.47$, $p = .157$, but there was a medium to large effect; $d = .70$.

While the difference between the groups was not significant, the experimental group did show significant gains (average 77.4% gain), from pre- to post-measure of multi-word expression use in their conversations $t(14) = -2.19$, $p = .046$, $d = .81$, 95% CI [-9.50, -.10]. By way of comparison, when the item scores for each expression in the multi-word expression cloze test were averaged and compared from pre- to post-test for only the experimental participants whose speech was analysed, there was a 30% gain from pre- to post-test knowledge of the target multi-word expressions.

4.3.1 Multi-word expression use gain scores

The change in use of multi-word expressions was also analysed through gain scores, whereby the increase in use of multi-word expressions between pre- and post-measure dialogue was calculated for each dyad. The distribution of these gain scores was normal for the control group but dubious for the experimental group with a significant Shapiro-Wilk score. Using an independent samples *t*-test, the average increases in multi-word expression use were compared between the experimental group ($n = 15$) $M = 4.80$, $SD = 8.48$ and the control group ($n = 8$) $M = 1.88$, $SD = 5.51$. The difference between the groups for change in multi-word expression use was not

significant $t(21) = -.877, p = .391, 95\% \text{ CI}[-9.86, 4.01], d = .408$. This analysis was supplemented with a Mann-Whitney test which confirmed there was no significant difference between change in use of multi-word expressions between the experimental group ($Mdn = 4.0$) and the control group ($Mdn = 2.0$) $U = 69.0, \text{ asymp } p = .559, z = .585, r = .122$.

4.3.2 Comparing MWE use with Wood (2009)

Wood (2009) found use of target multi-word expressions (those used in workshop models) increased from 2 to 18 after the fluency workshop treatment. The average use of multi-word expressions in the pre- and post-measure dialogues is shown by group and compared with the results reported by Wood (2009) in Figure 4.1. While the target multi-word expressions were not the same as in Wood's study, it is still interesting to compare the change in multi-word expression use after the fluency workshop in each study. The result from the present study which presents averages from across groups of participants, while showing the same trend of increase as in Wood (2009), is much less remarkable. The participant in Wood (2009) increased use of multi-word expressions nine-fold, whereas in this study the participants' increased use of target multi-word expressions was less than double.

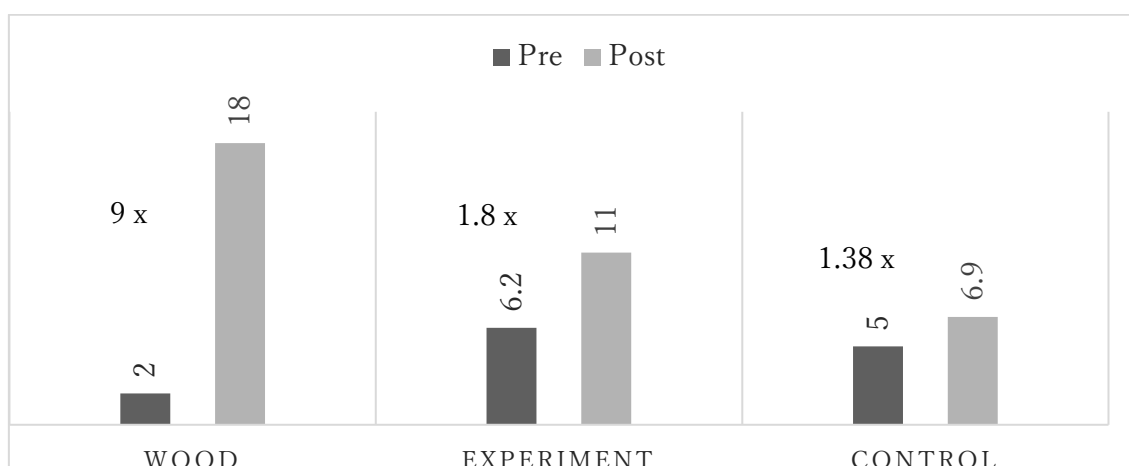


Figure 4.1 Average use of multi-word expressions by group

4.3.3 Comparing length of multi-word expressions

The mean length of multi-word expression was calculated by adding up the words used in expressions and dividing that word total by the total number of expressions used by that dyad to give an average length of multi-word expression. The differences between the groups and within the groups for mean length of multi-word expressions used in the dialogues were investigated. The small sample size (15 + 8) gave way to a non-normal distribution with a significant result for Kolmogorov-Smirnov and Shapiro-Wilk tests for both groups on pre- and post-measure data. Therefore, non-parametric tests were also used. For the pre-measure, a non-parametric Mann-Whitney test found no significant difference between the average length of multi-word expressions used by the experimental group ($Mdn = 2.00$, $Range = .500$) and the control group ($Mdn = 2.00$, $Range = 2.25$) $U = 71$, $asympt\ p = .282$, $z = 1.08$, $r = .224$. Therefore, we can assume the groups were able to use similar lengths of multi-word expressions before the treatment. For the post-measure, an independent t -test showed a difference between the mean length of multi-word expressions used by the experimental group and the control group, $t(21) = -2.15$, $p = .044$, $d = 0.86$. This analysis was supplemented with a Mann-Whitney test which showed a difference approaching significance ($p < 1$) with a medium effect size between the experimental group ($Mdn = 2$, $Range = 1.22$) and control group ($Mdn = 2$, $Range = 2.20$) $U = 83.0$, $asympt\ p = .081$, $z = 1.75$, $r = .364$.

4.3.4 Multi-word expression length gain scores

The gain scores or difference between average length of multi-word expressions in the pre- and post-measure dialogue were also calculated for each dyad and then subjected to an independent samples *t*-test to compare differences between the experimental group ($n = 15$) $M = .25$, $Mdn = .00$, $SD = .42$ and the control group ($n = 8$) $M = 0.24$, $Mdn = .00$, $SD = .71$. However, the difference between groups was not significant with a small effect size $t(21) = .038$, $p = .970$, $d = .017$. This result was confirmed using the non-parametric Mann-Whitney test $U = 78.0$, $asympt\ p = .172$, $z = 1.37$, $r = .285$.

4.3.5 Comparing multi-word expression length with Wood (2009)

Wood (2009) also counted the length of expressions used. The length of target multi-word expressions was capped at four words for this study, whereas in Wood (2009), there were 87 expressions ranging from two words to seven words, according to a file from Wood (personal communication, Sep 24, 2016). Wood (2009) found that the mean length of target multi-word expressions increased by 40.7% after his original fluency workshop. In comparison, the mean length of multi-word expressions increased by only 12.6%, $t(14) = 2.32$, $p = .036$, $d = .819$ ($Z = 21.0$, $p = .028$, $r = .402$) for the experimental group in this study. Meanwhile, the control group increased their mean length of multi-word expression by 16.3%, $t(7) = .971$, $p = .364$, $d = .297$ ($Z = 2.00$, $p = .045$, $st. z = 1.12$, $r = .280$). The small sample size and non-normal distribution should be noted when considering these percentages.

4.3.6 Extended multi-word expressions

Participants also used the target multi-word expressions in creative combinations, which were accurate but different from the target. Word combinations that were variations on a target multi-word expression and used to communicate a similar meaning to a target form were identified as an *extended multi-word expression* since they were extensions or variants on the targeted multi-word expression. For example, *would you like to* was a target multi-word expression. The combination *you like* was counted as multi-word expression use, while the combination *do you like* was counted

as an extended multi-word expression since *do* was not part of the target expression. These variants suggest familiarity with the target multi-word expression and could be considered as evidence of ability to use the expression. The use of variants could also signal partial ability to use the target expression. A list of the extended multi-word expressions identified in the transcripts is provided in Appendix 19.

4.3.7 Difference in length between MWEs and EMWEs

Table 4.3 shows the comparative lengths of multi-word expressions and extended multi-word expressions used in the dialogue for the experimental and control groups. Notice that there is only a slight increase in overall average length when the multi-word expression length is extended, on average in the pre-measure (experimental and control combined) from $M = 1.86$, $SD = .597$ to $M = 2.18$, $SD = .804$, however as this difference is statistically significant with a medium effect size $t(22) = 3.73$, $p = .001$, $d = .45$, it suggests perhaps that the counting of extended multi-word expressions was meaningfully different from the counting of multi-word expressions.

Table 4.3 Mean length of expressions by group

	Control ($n = 8$)	Experimental ($n = 15$) (difference with control)
Pre- length of MWEs	$M = 1.53$, $SD = .949$	$M = 2.03$, $SD = .129$ (+.50)
Post- length of MWEs	$M = 1.78$, $SD = .721$	$M = 2.29$, $SD = .429$ (+.49)
Pre- length EMWEs	$M = 1.75$, $SD = 1.13$	$M = 2.40$, $SD = .462$ (+.65)
Post- length EMWEs	$M = 2.18$, $SD = 1.02$	$M = 2.72$, $SD = .442$ (+.54)

Multi-word expressions (MWEs), extended multi-word expressions (EMWEs)

4.3.8 Length of extended multi-word expressions

An analysis of the mean length of extended multi-word expressions (EMWE) was also carried out to see if an increase in length of expression could be detected through extending the scope of the analysis. The pre-test length of extended multi-word expressions showed a non-normal distribution for both groups. The post-test was leaning away from normal for the control group with a couple of outliers but appeared reasonably normal for the experimental group. There were some ties in the data, that is,

some of the scores were the same between cases, so the asymptotic p -value is reported. An independent-samples Mann-Whitney test showed no significant difference between mean length of extended multi-word expressions in the pre-measure between the experimental group ($Mdn = 2.25$) and the control group ($Mdn = 2.00$) $U = 79.0$, asymp $p = .196$, $z = 1.29$, $r = .270$, and again no significant difference but a medium effect size in the post-measure between the experimental group ($Mdn = 2.63$) and the control group ($Mdn = 2.35$) $U = 86.5$, asymp $p = .085$, $z = 1.72$, $r = .359$. Therefore, the average length of extended multi-word expressions does not appear to be greatly different between the groups. When within group comparisons were made, a significant difference was found within the experimental group from pre-intervention mean length of extended multi-word expression to post-measure $t(14) = 2.77$, $p = .015$, $d = .711$. This difference was confirmed with a non-parametric version of the paired samples t -test, the Related-Samples Wilcoxon signed rank test, ($Z = 79.0$, asymp $p = .019$, $z = 2.34$, $r = .428$). There was no significant difference within the control group between the pre- mean length of extended multi-word expression and post-measure mean length of extended multi-word expression, $t(7) = 1.14$, $p = .293$, $d = .399$, and this result was confirmed with Wilcoxon (since normality was in doubt) ($Z = 12.0$, asymp $p = .225$, st. $z = 1.21$, $r = .304$). These results suggest that the experimental group were able to increase the length of EMWE they used, but this increase was not more than the control group. The results for this extended analysis did not differ dramatically from the original multi-word expression length analysis as they also showed no significant difference in length.

4.3.9 Answering RQ. 2

Research question two asked: *Does the experimental group use more multi-word expressions in speech than the control group?* This question was investigated through analysing the change in multi-word expression use and length. The results from this study show no statistically significant difference in use between the groups, although there was a medium to large effect size which indicates a difference. There was also a difference approaching significance with a medium effect size in the average length of multi-word expressions used between the experimental participants and the control group. Extending the scope of the multi-word expressions to include slight variation

showed the same comparison patterns with greater difference between the groups as the target multi-word expression analysis. These results suggest that there may be some advantage from the fluency workshop intervention for increasing productive use of multi-word expressions. The next section reports the measures of spoken fluency.

4.4 Spoken fluency

In order to assess changes in spoken fluency, a spoken role-play dialogue by participants was audio recorded before and after the treatment (for details see 3.7.3). Only dialogue recordings which had the same speaking partners in both the pre- and post-measure were analysed. This criterion resulted in 23 sets of useable role-play dialogues. The following results are based on analysis of the dialogue transcriptions and audio files.

Fluency was measured by speed of delivery; that is total meaningful syllables uttered per minute (trimmed speech rate). Fillers and repetitions are an important and natural aspect of conversation so untrimmed measures of speech rate and mean length of runs (uninterrupted utterances between pauses) were also calculated and compared. For more discussion on fluency measures, see 5.7 and 6.5. Table 4.4 shows the trimmed speech rate averages before and after the treatment.

Table 4.4 Trimmed speech rate in the pre- and post-test

Trimmed speech rate	Pre-test mean	<i>SD</i>	95% <i>CI</i>	Post-test mean	<i>SD</i>	95% <i>CI</i>
Experimental (<i>n</i> = 15)	47.0	16.4	37.2, 56.9	58.0*	18.5	49.1, 66.8
Control (<i>n</i> = 8)	47.7	21.6	34.3, 61.2	53.6	11.3	41.5, 65.7

* indicates significant difference at .025 (Bonferroni corrected) between pre- and post-test mean (within group)

An independent samples *t*-test showed that there was no significant difference between the trimmed pre-test speech rate of the experimental group and the control

group; $t(21) = .088, p = .930, d = .036$. There was also no significant difference for the post-test trimmed speech rate between the experimental and the control groups; $t(21) = .608, p = .550, d = .287$. A similar result was found for the untrimmed speech rate comparisons between groups. However, when considered alone, the experimental group increased their trimmed speech rate significantly $t(14) = -3.64, 95\% \text{ CI } [-19.83, -5.12], p = .003, d = .68$, while the control group did not $t(7) = -.55, p = .60, 95\% \text{ CI } [-20.61, 12.85], d = .21$.

4.4.1 Speech rate gain scores

Fluency changes were double-checked by calculating gain scores (subtracting the pre-test speech rate from the post-test speech rate). An independent samples t -test showed that there was no significant difference between gain scores in trimmed speech rate between the experimental group ($n = 15$), ($M = 10.9, SD = 12.9$) and the control group ($n = 8$), ($M = 5.84, SD = 23.6$), $t(21) = .674, p = .508, d = .266$. However, a medium effect size was evident when the untrimmed speech rate gain scores were compared between the experimental group ($M = 12.47, SD = 13.28$) and the control group ($M = 3.87, SD = 20.01$), $t(21) = 1.3, p = .229, d = .51$. The untrimmed speech rates are shown in Table 4.5, with the percentage gain from pre- to post-measure in brackets under the post speech rate.

Table 4.5 Descriptive statistics for pre- and post-measure untrimmed speech rate

	Untrimmed pre speech rate			Untrimmed post speech rate		
	M	SD	95% CI	M	SD	95% CI
				(difference from pre)		
Experimental ($n = 15$)	50.82	18.13	40.78, 60.86	63.29	18.54	53.02, 73.56
				(+24.5%)*		
Control ($n = 8$)	53.23	23.43	33.65, 72.82	57.11	12.34	46.80, 67.42
				(+7.3%)		
Wood ($n = 1$)	123			140		
				(+13.8%)		

4.4.2 Speech rate compared with Wood (2009)

In Wood (2009), where there was one participant, speech rate increased by 13.8% $(\text{post-test rate} - \text{pre-test rate}) / \text{pre-test rate} \times 100$) from pre- to post-test measure. There is no mention in the research results of trimming the speech rate, so we must assume that the speech rate is untrimmed. To compare, in the present study, the experimental group's untrimmed speech rate increased by 24.5%, a significant increase according to a paired samples t -test; $t(14) = -3.64$, 95% CI $[-19.8, -5.1]$, $p = .003$, $d = 0.68$, and the control group increased by 7.3%; $t(7) = -.548$, 95% CI $[-20.6, 12.9]$, $p = .60$, $d = 0.21$. The results from the present study support Wood's result; that is, the speech rate increased for the experimental group. Figure 4.2 shows the increase in speech rate for the Wood (2009) study alongside the increase in speech rate for the present study. It is obvious from the figure that the speech rate was greater to begin with in the Wood (2009) study, which could reflect the differences between the two study designs; the monologue vs. dialogue format and the English language proficiency differences between the participants. The greater proportional increase in speech rate for the experimental participants in this study may also be a reflection of these differences.

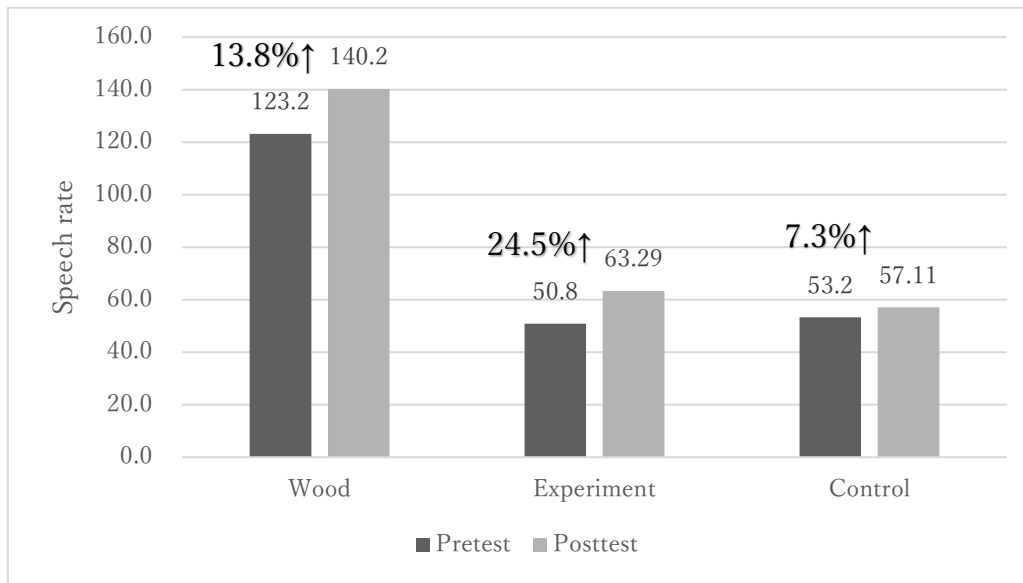


Figure 4.2 Speech rate in the pre- and post-measure compared with Wood (2009)

4.4.3 Mean length of runs

The second measure of fluency was mean length of runs. The mean length of runs was calculated as total number of syllables uttered divided by the number of runs. Runs were demarcated by change of speaker and audible pauses between runs of speech. No significant difference was found between the mean length of runs for the experimental group ($n = 15$) $M = 3.78$, $SD = .547$ and the control group ($n = 8$) $M = 3.86$, $SD = .613$; $t(21) = .320$, $p = .752$, $d = .138$ in the pre-measure. There was also no significant difference for mean length of runs in the post-measure between the experimental group ($n = 15$) $M = 3.62$, $SD = .620$ and the control group ($n = 8$) $M = 3.58$, $SD = .402$; $t(21) = .216$, $p = .831$, $d = .077$. Table 4.6 shows the descriptive statistics for the mean length of runs for both trimmed (trim) and untrimmed (untrim) audio recordings, with the difference between the pre-and post-measure shown in parentheses.

Table 4.6 Descriptive statistics for pre- and post-measure mean length of runs

	Pre- mean length of runs			Post- mean length of runs		
	Trim	Untrim	Trim to untrim difference	Trim (difference from pre)	Untrim (difference from pre)	Trim to untrim difference
Experimental ($n = 15$)	3.78	3.70	-.08	3.62 (-4.2%)	3.53 (-4.6%)	-.09
Control ($n = 8$)	3.86	4.13	+.27	3.58 (-7.3%)	3.35 (-18.9%)	-.23
Wood ($n = 1$)		5.1			6.4 (+25.5%)	

4.4.4 Mean length of runs pre- to post-measure gain scores

The difference between pre- and post-measure mean length of runs was calculated for each dyad, and then the average differences compared across the groups. The distribution of pre-post differences in the experimental group included an outlier, with a significant Kolmogorov-Smirnov score; $D(23) = .190$, $p = .030$. An independent samples t -test showed no significant difference for mean length of runs from pre- to post-measure between the experimental group ($n = 15$) $M = .156$, $Mdn = -.212$, $SD = .554$ and the control group ($n = 8$) $M = -.288$, $Mdn = -.302$, $SD = .370$, $t(21) = -.606$, $p = .551$, $d = .280$. This finding was supplemented and corroborated with a Mann-Whitney U test ($U = 66.0$, asymptotic $p = .699$, $z = .387$, $r = .081$) which also showed only a small effect size. The gain scores for the untrimmed mean length of runs were also not significantly different between the groups ($U = 89.0$, asymptotic $p = .061$, $z = 1.87$, $r = 0.39$) but did show a medium effect size.

4.4.5 Answering RQ3

Research question three asked: Does the experimental group experience greater increase in spoken fluency than the control group? Spoken fluency was measured through speech rate and mean length of runs. Results from this study show that although there was a significant increase in speech rate for the experimental group, it was not significantly greater than the increase in speech rate experienced by participants in the control group. Also, the mean length of runs did not change significantly after the treatment between the groups. Wood (2009) found an increase in mean length of runs after the treatment for a monologue narrative. In the current study, mean length of runs appeared to decrease in the second sample, suggesting that there might be a difference here depending on format (monologue or dialogue) as to whether mean length of runs can be expected to increase. It should be noted that for dialogue Tavakoli (2016) found that the mean length of runs was actually longer when compared with monologue samples from the same speakers. Rapid turn-taking in a dialogue format could also be an indication of interactive speech fluency and may actually shorten the mean length of runs. If this is the case, then increased ability to produce longer runs may not manifest itself because it is offset by an increased interactional fluency. It could also be that certain types of dialogue are more likely to draw out longer runs than others. These results suggest that the fluency workshop has potential to improve speech rate, but not necessarily to lengthen uninterrupted runs of speech. It could also be that mean length of runs may not be the best measure to judge conversational fluency with.

Since the mean length of runs decreased in the post-measure, and there was a suggestion that shorter turn taking could be the cause, the number of turns (runs) per minute was investigated. Conversation turns per minute were calculated by dividing runs by total seconds and multiplying by 60. Table 4.7 shows that there was actually a gain (shown in parentheses) from pre- to post-measure for both groups for runs per minute. As the untrimmed data contained all interactions, that is what is reported here. When the groups were combined, there was a significant increase in runs per minute from pre- ($M = 13.72$, $SD = 5.30$) to post- ($M = 18.13$, $SD = 5.86$) measure, $t(22) = -4.97$, $p < .01$, $d = .79$, with a large effect size. It appears that the mean length of runs was unchanged, although the runs per minute increased in both groups, suggesting that

they were taking more frequent short conversational turns. There was a potentially negative correlation found between the gain scores for runs per minute and mean length of runs ($n = 23$), $r = -.203$, 95% CI $[-.29, .57]$, however the confidence intervals pass through zero which raises the possibility of zero correlation. There was no significant difference found between the gain scores for runs per minute from pre- to post- measure between the experimental group ($n = 15$) $M = 4.62$, $SD = 3.85$ and the control group ($n = 8$) $M = 4.01$, $SD = 5.17$, $t(21) = -.322$, $p = .75$, $d = .13$. The discussion in Chapter Seven will consider the influence of the experimental activities on speaking fluency further.

Table 4.7 Descriptive statistics for pre- and post-measure runs per minute

	Runs per minute pre-measure		Runs per minute post-measure	
	Trim	Untrim	Trim (difference from pre)	Untrim (difference from pre)
Experimental ($n = 15$)	12.47	13.87	16.17 (+29.7%)	18.49 (+33.3%)
Control ($n = 8$)	12.64	13.45	15.03 (+18.9%)	17.46 (+29.8%)

4.5 Relationship between fluency and use of multi-word expressions

This next section investigates whether a relationship could be seen between fluency and use of multi-word expressions. Correlations between the various fluency measures and multi-word expressions are reported.

4.5.1 Correlation between speech rate and multi-word expression use

It was hypothesized that trimmed speech rate (fluency) and use of multi-word expressions would correlate positively. Pearson's correlation test shows the strength of correlation from 0 to 1 (0 - 100%) correlation, where r is the estimate of the correlation and the confidence intervals show the range in which the actual correlation lies. The following correlations have been done with 1000x bootstrapping; a "nonparametric

approach to statistical inference” done by computer which randomly resamples with replacement from the sample many times in order to be able to make statistical inferences about the data set (Lavrakas, 2008, p. 65). In the pre-test, trimmed speech rate and use of multi-word expressions appeared to have a positive correlation ($n = 23$), $r = .484$, 95% CI [.140, .759]. However, in the post-test, the confidence intervals pass through zero ($n = 23$), $r = .237$, 95% CI [-.86, .608], suggesting the possibility of zero correlation. When the gain in trimmed speech rate was related to the gain in the use of multi-word expressions (regardless of group), as seen in the scatterplot in Figure 4.3, the correlation confidence intervals also passed through zero ($n = 23$), $r = .311$, 95% CI [-.032, .664]. These results suggest a positive correlation is possible between trimmed speech rate and use of multi-word expressions, but certainly do not prove it.

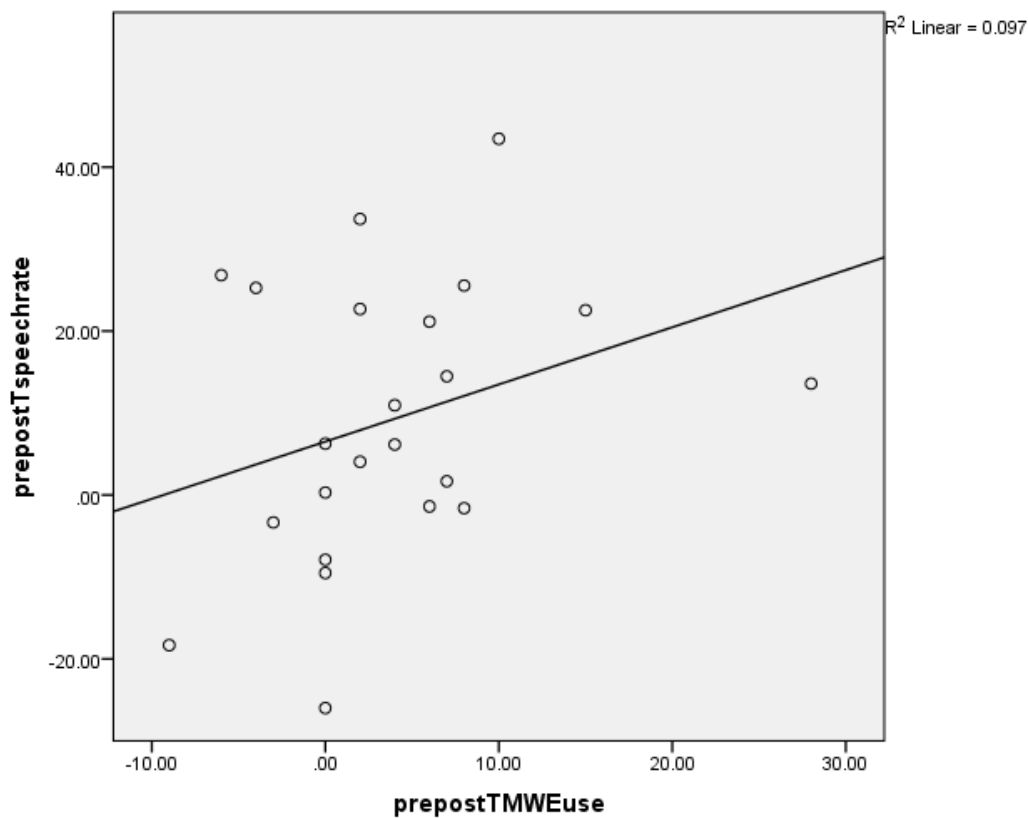


Figure 4.3 Correlation between change in speech rate and change in multi-word expression use

4.5.2 Correlation between gains in speech rate and gains in MWE length

It was also hypothesized that fluency (trimmed speech rate) would increase with length of multi-word expressions. The correlation between trimmed speech rate gain scores and length of multi-word expression gain scores ($n = 23$), $r = .370$, 95% CI $[-.136, .665]$ was again not clear. The relationship can be seen in the scatterplot shown in Figure 4.4 where the majority of data points are not following the correlation line.

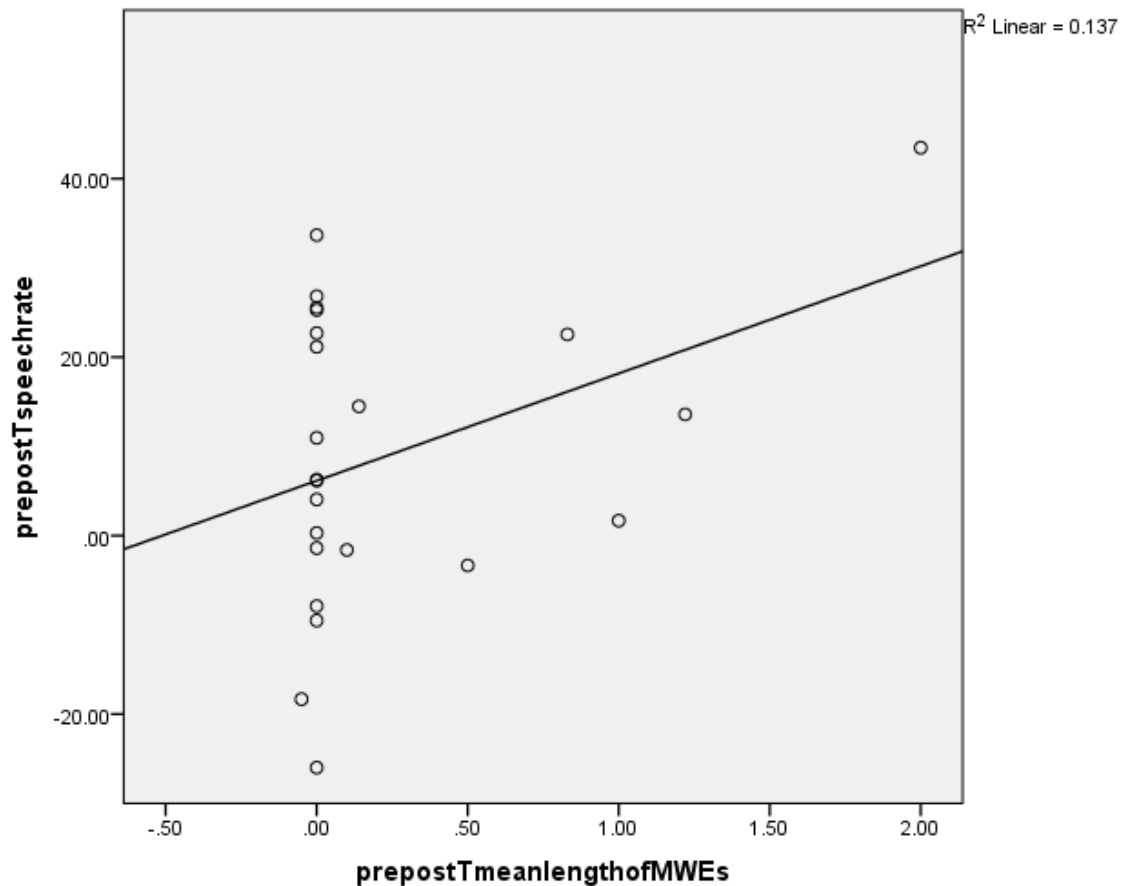


Figure 4.4 Correlation between change in speech rate and change in length of multi-word expressions

4.5.3 Correlation between gains in mean length of runs and MWE use

The relationship between mean length of runs (another measure of fluency) and use of multi-word expressions was also of interest. It was hypothesized that the mean length of runs would increase as the use of multi-word expressions increased. This

hypothesis was investigated through a Pearson's correlation of gain scores for mean length of runs and gain scores for multi-word expression use. Since the confidence intervals go through zero, the strength of any correlation is in doubt ($n = 23$), $r = .518$, 95% CI $[-.172, .792]$. The correlation is illustrated in the scatter plot in Figure 4.5, where the spread of data points is not tightly following the correlation line.

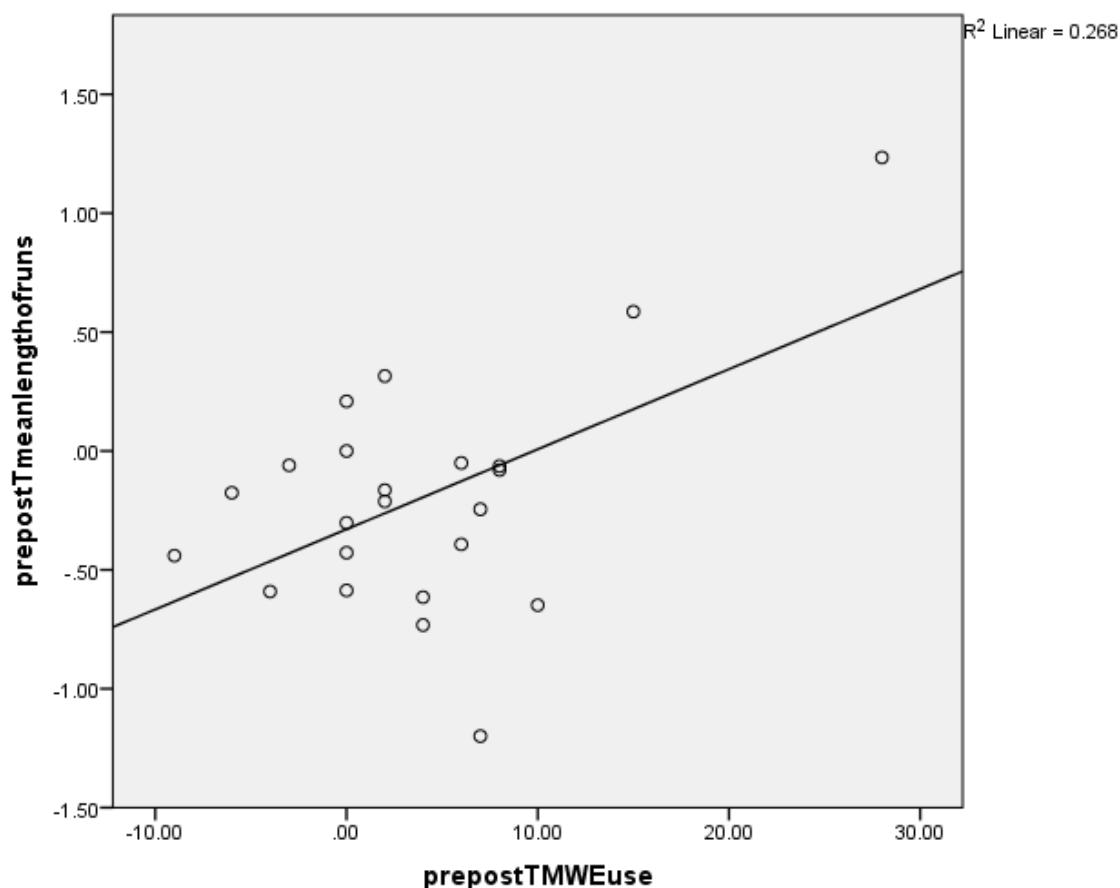


Figure 4.5 Mean length of run gain correlated with multi-word expression use gain

4.5.4 Answering RQ. 4

Returning to research question four, is there a relationship between fluency and use of multi-word expressions? These results fail to provide any strong evidence of such a relationship, though they are perhaps suggestive of a positive correlation. This research question will be discussed further in Chapter Seven. The next section will investigate whether the experimental intervention had any impact on general vocabulary learning.

4.6 Aural vocabulary knowledge

This next section investigates changes in vocabulary knowledge from before and after the intervention in order to discover if a focus on learning multi-word expressions limits wider vocabulary learning.

4.6.1 Listening Vocabulary Levels Test (LVLT)

Participants completed the first three levels of the Listening Vocabulary Levels Test (LVLT) (McLean et al., 2015) before and after the treatment, information on the test can be found in Chapter Three. Test answers were electronically recorded via Qualtrics and then exported and analysed using SPSS to check for differences between the two groups. Looking at the overall scores for this test, one extreme outlier was identified and removed by excluding those scoring less than 25 out of 72 on this test. A score of less than 25 (less than half of the average score of the participants) was interpreted as indicative that a participant was not paying full attention to the test. I had also noticed that this particular participant was not giving the test his full attention. Descriptive statistics for the average test scores for pre- and post-test can be seen in Table 4.8. Notice there is not a lot of difference between the scores, and both groups increased in the post-test by a similar amount.

Table 4.8 Descriptive statistics for LVLT results across groups

	<i>n</i>	Pre-test <i>M</i>	95% <i>CI</i>	<i>SD</i>		<i>n</i>	Post-test <i>M</i>	95% <i>CI</i>	<i>SD</i>
Control	29	57.97	55.40, 60.53	6.75		29	58.93	56.62, 61.24	6.07
Experiment	44	56.91	55.22, 58.59	5.54		44	57.89	56.13, 59.64	5.77

4.6.2 Pre-test vocabulary scores

An independent *t*-test was used on the pre-test scores and showed no significant difference between the experimental and control groups, $t(71) = .730$, $p = .468$, $d = .171$.

The small effect size further indicates that the group averages were not meaningfully different.

4.6.3 Post-test vocabulary scores

Post-test score distributions showed deviations from normal with significant scores in the Kolmogorov-Smirnov and Shapiro-Wilk normality tests. An independent *t*-test showed no significant difference between the experimental group (*Mdn* = 58, *Range* = 21) and the control group (*Mdn* = 58, *Range* = 21), $t(71) = .742, p = .461, d = .175$ in the post-test scores. A Mann-Whitney test confirmed this result since the data was not normally distributed ($U=558, z = .910, \text{asympt } p = .363, r = .107$).

4.6.4 Vocabulary gain scores

Gain scores were calculated from pre- to post-test. The experimental group showed non-normal deviation with significant Kolmogorov-Smirnov and Shapiro-Wilk scores. An independent *t*-test on the gain scores found no difference between the experimental group ($M = .977, SD = 5.80$) and the control group ($M = .966, SD = 3.02$) $t(71) = .010, p = .992, d = .002$. This result was confirmed with a Mann-Whitney test ($U = 629.0, z = .102, \text{asympt } p = .919, r = .012$), since normality was in question.

4.6.5 Vocabulary scores by level

The vocabulary scores were also analysed separately for each frequency level within each group to see if there were any within group gains in vocabulary knowledge. Tests of normality showed non-normal distributions for the first two levels of the test. Scoring 23 out of 24 on a test level could be interpreted as ability to comprehend texts made up of words from that frequency level (McLean et al., 2015, p. 14). Both groups had medians of 23 for level one (the most frequent 1000 words in English) but were below this threshold on the second (the most frequent 2000 words in English) and third (the most frequent 3000 words in English) levels. Table 4.9 shows the median scores and within group comparisons on each level. The only significant pre- to post-test change within the groups was an increase for the control group on level two (the 2000 most frequent words). The Shapiro-Wilk scores show the lack of normal distribution in the data.

Table 4.9 Listening Vocabulary Levels Test results by level and group

Experimental (<i>n</i> = 44)	<i>Shapiro-Wilk</i> (<i>sig</i>)	<i>Median</i>	<i>IQR</i>	<i>p</i> value for difference between pre- and post-test
Level 1 pre	.858 (< .001)	23.0	1.75	
Level 1 post	.838 (< .001)	23.0	2.0	.112
Level 2 pre	.933 (.013)	20.0	3.0	
Level 2 post	.813 (< .001)	21.0	3.0	.873
Level 3 pre	.974 (.414)	14.0	3.0	
Level 3 post	.954 (.078)	15.0	3.0	.147
<hr/>				
Control (<i>n</i> = 29)				
Level 1 pre	.741 (< .001)	23.0	2.0	
Level 1 post	.720 (< .001)	23.0	2.5	.289
Level 2 pre	.970 (.565)	20.0	3.0	
Level 2 post	.902 (.011)	21.0	2.0	.001*
Level 3 pre	.961 (.354)	16.0	5.0	
Level 3 post	.954 (.230)	15.0	6.0	.341

* $p < 0.05$

4.6.6 Answering RQ 5

Based on these results, it would appear that the focus on multi-word expressions in the experimental group did not impede general vocabulary learning any more or less when compared with the control group. For further discussion, see 7.2.2. Observations and perceptions from the classroom will be considered in the next section.

4.7 Participant feedback regarding experimental activities

In this next section I consider participant feedback through surveys (see Appendix 16 for the format example) and focus group interviews regarding the experimental classroom activities. I wanted to discover if adaptations to the fluency workshop activities were necessary for the dialogue format and EFL context. Therefore, I surveyed participants about their perceptions of the experimental class activities, asking

about the perceived usefulness and difficulty for each activity. Choosing five as a response on the Likert scale indicated strong agreement with the statement. Choosing numbers decreasing from five indicated decreasing levels of agreement with the statement. Three was the middle point of the Likert scale and represents neither agreeing nor disagreeing with the statement. There is evidence that Japanese prefer to choose middle points rather than choosing a positive response in such scales (Lee et al., 2002). There was certainly a tendency towards the middle score of three in these survey results which suggests either ambivalence or satisfaction. The results from this survey will be considered critically below, starting with those activities where issues were identified, and ending with those activities that were found to be satisfactory as they were. I aim to give a rich description of what happened in the classroom by including remarks from focus group interview participants and participant comments from the surveys, along with my own observations as the teacher/researcher regarding the classroom activities. The activity feedback survey results are also presented using bar charts. Focus group participants are referred to using pseudonym names to safeguard anonymity. In the first focus group, there was one Chinese female (Elaine) who had a higher English proficiency than most others in her class. In the second focus group, there were three Japanese females (Hana, Yumi and Yuki) who had a similar English proficiency to others in their class. Each section concludes with a decision about whether to keep the activity as is, make changes or remove it from the intervention for Study Two. The first activity to be considered was marking pauses.

4.7.1 Marking pauses

Marking pauses required students to listen to the audio dialogue and mark audible pauses with a back-slash mark in the transcript. Listening for and marking audible pauses was a new concept for most. As the teacher, I wondered whether this activity was relevant to this type of dialogue text because pauses were few since conversation turns were frequent. The focus groups and survey comments revealed confusion and dissatisfaction with the activity. In Figure 4.6, participant responses regarding the usefulness of marking pauses for learning English and whether it was too difficult as an activity are shown. For ease of viewing, responses have been amalgamated into three categories of response to the statement; disagree, neutral and agree. In the focus group,

Hana, Yumi and Yuki all chorused that, “the sound was too fast” and “mostly, to be honest you cannot hear any pauses.” Hana admitted to inserting pause marks where she thought there were meaning boundaries, or where she thought they should go rather than because she could hear a pause. Yumi also commented that marking pauses before learning the meaning seemed premature, “we were doing it before we understood the meaning.”

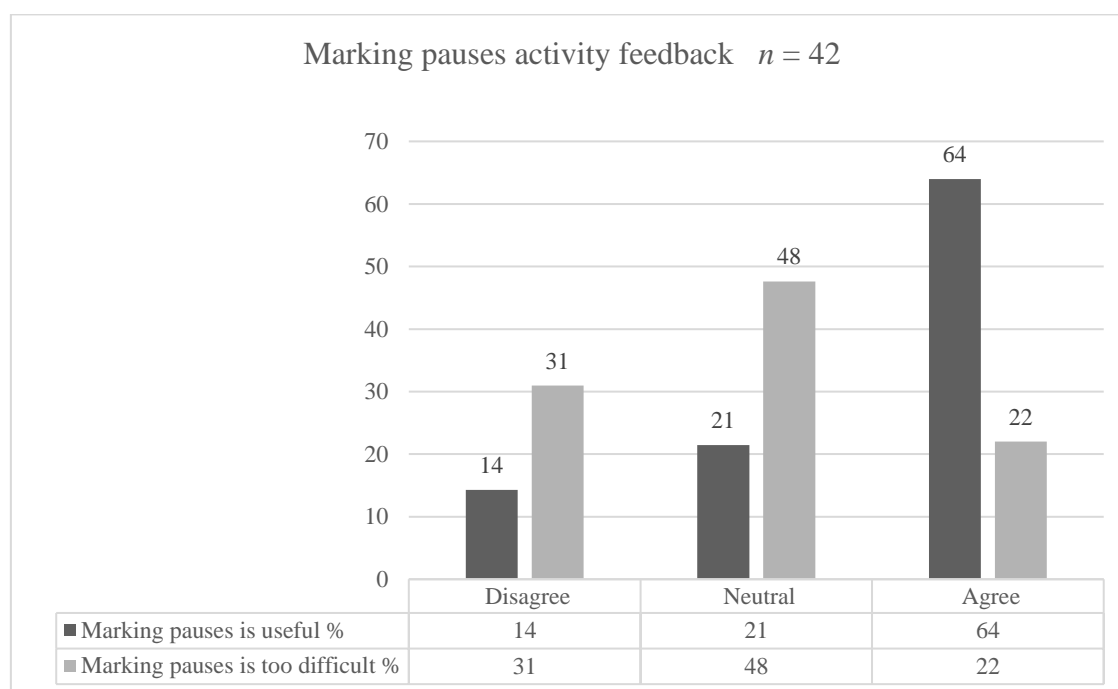


Figure 4.6 Activity feedback survey results for marking pauses

Regarding perceived usefulness, the majority 64% considered marking pauses a useful activity for learning English, it was also not considered a difficult activity with only 22% considering it too difficult. One comment from the survey stated, “...(I) could not sense the pauses, but marking pauses was useful for doing the role-play.” In Wood (2009), the model text was a narrative, presumably with more identifiable pauses. However, with the short-turn dialogue texts used in this class, audible pauses were few and the value of this activity considering participant feedback and my own observations was questionable. I decided therefore that the marking pauses activity was probably inappropriate for these short turn dialogues. In Chapter Five, I describe the decision to remove this activity and what it was replaced with. The next activity to investigate is the role-play recording.

4.7.2 Role-play recording

Role-play recording required students to audio record their role-plays without reference to notes, they then had to save and export their audio file. It was not a straightforward activity. As the teacher, I found the role-play recording activity the most complicated. There were always students who needed assistance with the technical side of the activity, and it would therefore frequently take around 20 minutes for a three-minute audio recording. Elaine mentioned that participants were often nervous when recording which made it easier to make mistakes, “if you are going to record you feel nervous, so it is easy to make mistakes.” Yumi noted, “we couldn’t refer to our notes, so it was difficult.” Hana and Yumi stated that they did not listen back to the recording, so they saw little point in the activity, “I didn’t feel like recording was very meaningful.” Yuki also noted that, “if there was an opportunity to listen to it, it would be useful.” Elaine talked about how she enjoyed listening back to the recording afterwards to spot her own mistakes, “if you record then you can listen to your own voice afterwards, if you made a mistake you can improve on it.” Yumi noted and Hana and Yuki agreed with her that the actual logistics of recording the role-play and then uploading it was complicated and confusing, and a distraction from focusing on the language, “after recording, you have to change it from MP3, that work is difficult.”

Figure 4.7 shows survey responses from participants. 73% of participants responded that they thought the role-play recording was a useful activity for learning English, and 28% considered the recording activity too difficult.

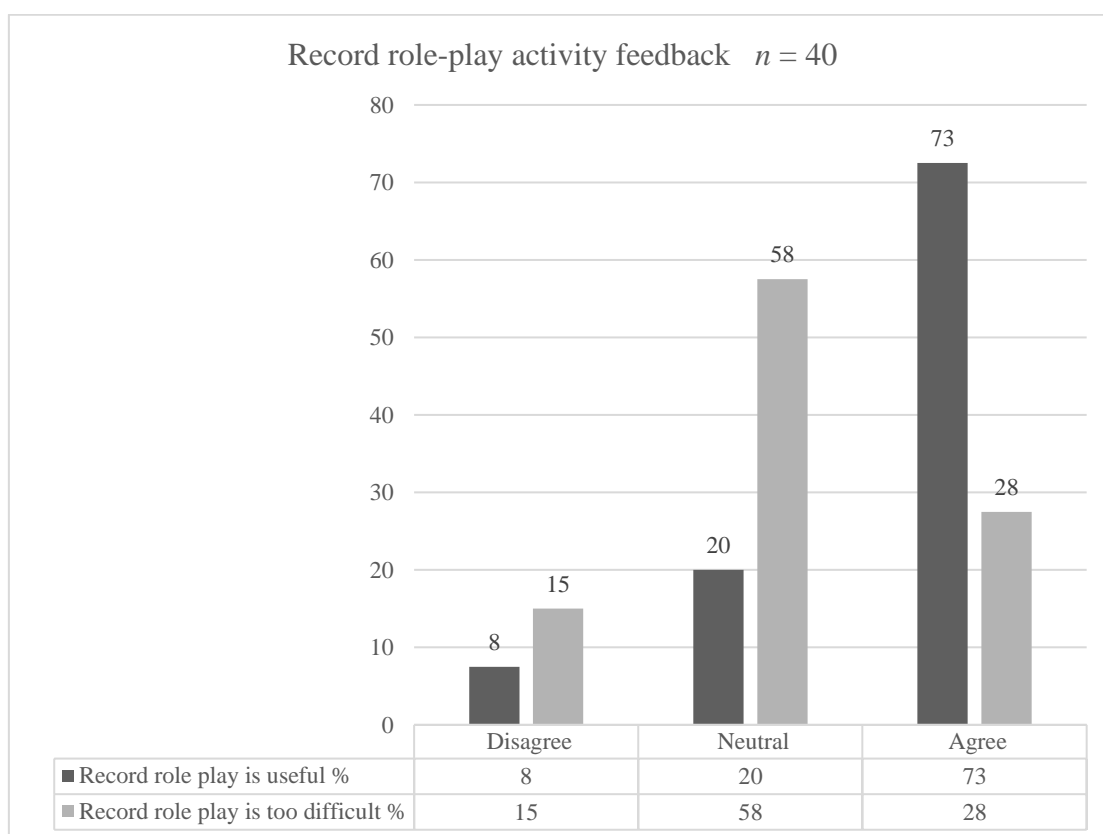


Figure 4.7 Activity feedback survey results for role-play recording

The focus group participants noted that recording the role-play was only useful if one listened back to it critically, and time was not given for this in class. Therefore, in order to improve the face validity of the recording activity and give learners the opportunity to reflect on their fluency development, a structured reflection time was added after each recording for Study Two, this adaption is described in more detail in Chapter Five. A time for reflection on the recording was included in Wood (2009), therefore bringing reflection into the class time would increase the similarity of Study Two with the original fluency workshop. The next activity to be considered is the dictogloss.

4.7.3 Dictogloss

Dictogloss required students to listen to the audio, take notes and try to reconstruct the dialogue in writing. Of the four class groups that participated in the experimental activities, there was one class that was smaller and appeared more comfortable with each other than the other classes. For the smaller class, the dictogloss

was a good activity, and within their ability. But in the other three classes, I observed most participants attempting to write the text word for word. Also, many were tempted to look back to their transcript to copy, which was discouraged but difficult to prevent completely. There were some students who worked well with their partner to reconstruct the text. However, some partners did not work well together, for example only showing their work to their partner when I nudged them. I also noted that students were slow to interact in the final group effort to complete the task. Additionally, Hana, Yumi and Yuki said the dictogloss audio playback was too fast. They suggested having shorter segments to write (more like a gap-fill) or pausing the audio, “I can hear it, but I can’t write it. The speed is too fast, if the audio was stopped in sections I would be able to write.” However, Elaine, who had higher English proficiency, enjoyed dictogloss because it provided both listening and writing practice, “because it is not just listening, it is also writing practice.” When the issue of copying was raised in the focus groups, all four participants thought this was probably because the students felt the activity was too difficult. Regarding writing word for word, Hana and Yumi said that they understood they could write the meaning in a different way, but said they did not have the language, hence the desire to write the exact words, “I can’t come up with the same meaning in a different way.” Figure 4.8 shows 54% considered dictogloss to be a useful activity, but 64% considered it to be too difficult, which is backed up with the participant comments.

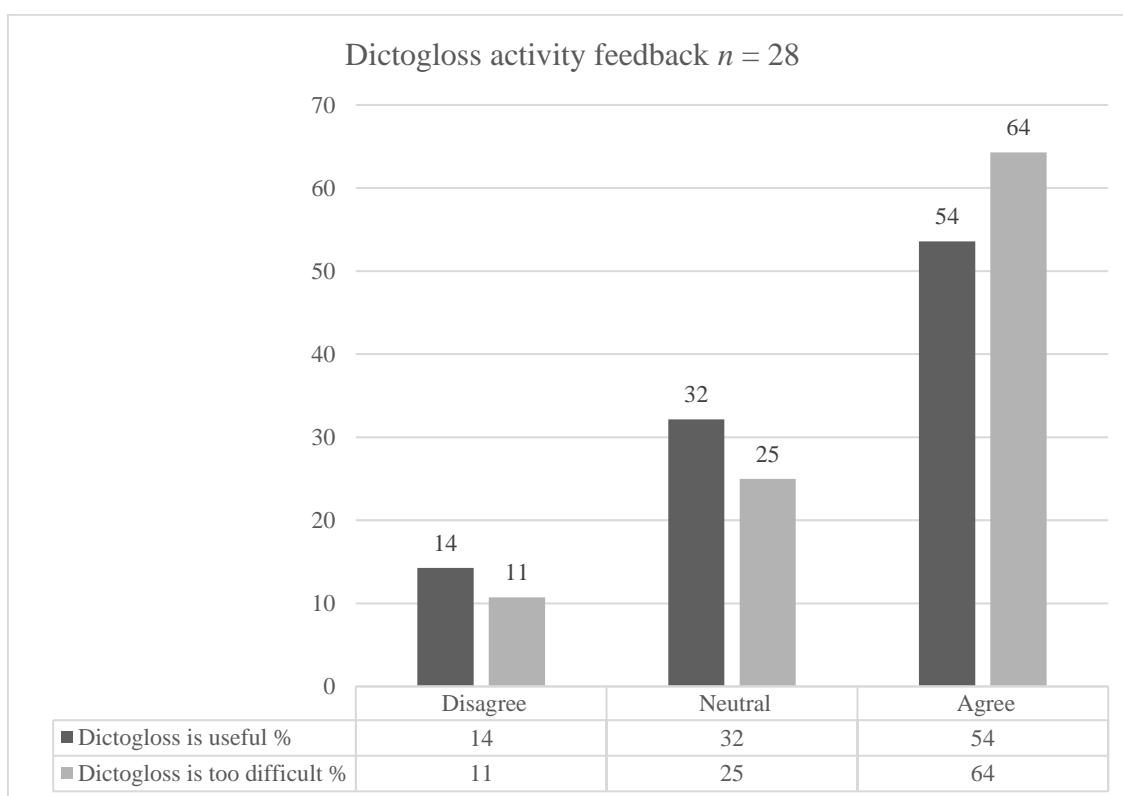


Figure 4.8 Activity feedback survey results for dictogloss

In order to provide more support for the dictogloss in Study Two, I decided to have students work in groups from the beginning to increase their chances of supporting each other and impose a time limit on their interaction to motivate them to start interacting with less hesitation. I also decided to add more structure to the print handout to help students keep track with the audio, further detail regarding these adaptations is provided in Chapter Five. The next activity to be considered is the mingle jigsaw.

4.7.4 Mingle jigsaw

The mingle jigsaw required participants to walk around the classroom sharing memorised multi-word expressions with each other. I thought that the mingle jigsaw would be one of the most popular activities because it allowed students to stand up and walk around talking to people of their choice. However, mingling in most of the classes was a bit lethargic, which surprised me. The mingle jigsaw seemed less a linguistic challenge and more a social communicative challenge; whether one could approach another and ask them about their phrase. Yumi and Hana remarked “it is more about

communication ability than English. I am not sure if I can ask people and friends what is the phrase you remember.” In this particular context, an engineering university with most students being male, there was a marked lack of openness to social interaction, with some boys not moving around to exchange phrases and not looking for interaction. Yumi complained, “many of the boys really do not move themselves.” Hana, Yumi and Yuki noted that exchanging phrases was easy with people they were friendly with, but it became difficult and awkward to approach people who seemed uninterested in communicating. This social awkwardness would make them give up collecting all ten phrases in order to avoid risk in social interaction. Yumi explained, “I can ask friends, but is this person okay to ask? I end up not able to ask, I can’t collect them all, at a certain point I give up.” Elaine mentioned that if the mingle jigsaw phrases were too short, it made them hard to understand, “if the phrase is too short, I can’t really understand the meaning.” Figure 4.9 shows that 52% of participants thought the mingle jigsaw activity was useful for practising English, while the majority had no issue with the difficulty level.

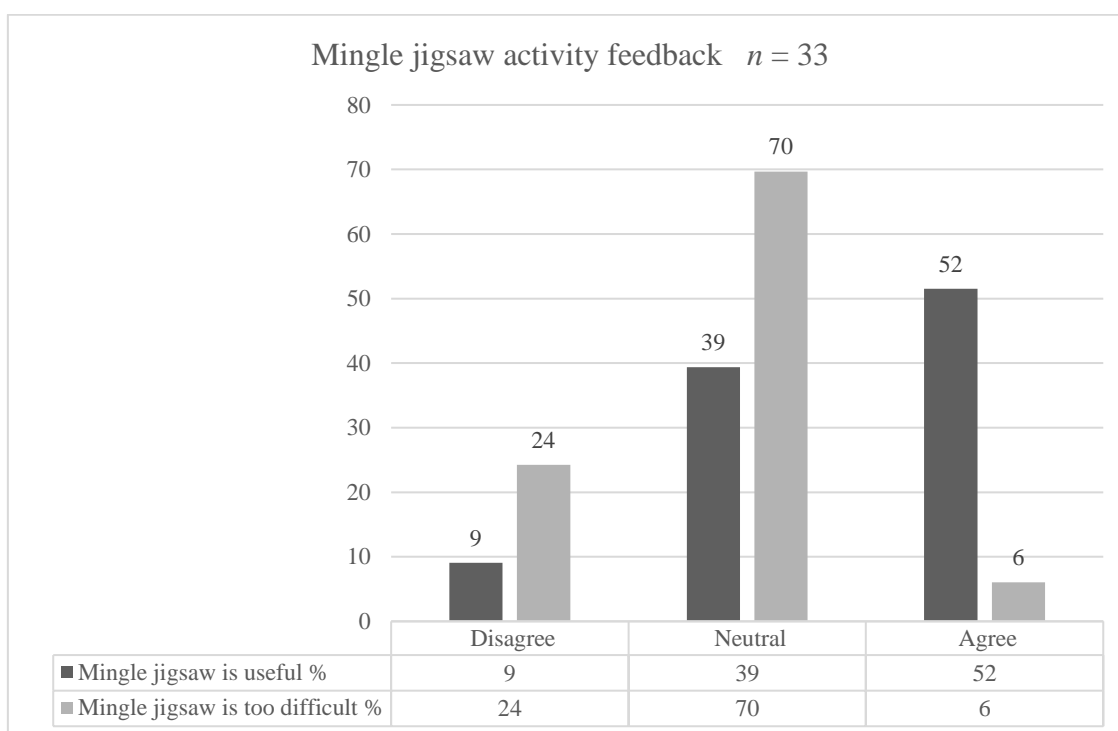


Figure 4.9 Activity feedback survey results for mingle jigsaw

Considering my observations and participants' comments, I decided to put a visible five-minute time limit for mingling and sharing multi-word expressions in the hope that limiting the time would increase the pressure and energy for this activity. This adaptation is explained more in Chapter Five. The next activity to consider is the decreasing time role-play.

4.7.5 Decreasing time role-play

The decreasing time role-play required students to role-play without their notes under increasing time pressure. As the teacher, I enjoyed the added excitement of the time pressure, and really wanted students to be able to role-play without their notes. Students, however, were apt to glance at notes if they were in reach, and shortcut the dialogues on occasion, especially if it appeared that others had finished talking. Elaine said that the increasing time pressure helped her to remember the conversation, "as the time gets shorter, my head can work faster, if it is like that then I can remember that conversation." It was noted by Yumi that without notes, there was a risk of the conversation stopping when someone could not remember what to say, so they liked to

have the notes within reach for such moments, “without notes, I can’t do it. If my partner also can’t do it, the conversation cannot progress. If I can glance at the notes I can remember.” But their reliance on notes depended on the scenario, with the directions scenario being the easiest. The hotel scenario had a more checklist like set of questions to cover and was therefore more difficult without notes, particularly for the person playing the hotel staff role. Hana explained, “for directions, I can do it without paper somehow, but the hotel staff role is the most difficult, you have to do it the proper way.”

Hana suggested that two minutes was too long for the restaurant and directions scenarios, with the final 30 seconds falling into silence, “with 2 minutes, the remaining 30 seconds turn into silence.” In fact, it was suggested that the shorter time limits were preferred. Yumi said she deliberately spoke slowly during the first longer time periods in order to avoid vacant time, “that’s why I speak very slowly.” However, Yuki noted that the same time allocation for the hotel role-play was too short, they all agreed with this, “for the café and directions there is left over time, but the hotel you really only get halfway through.” Hana and Yumi suggested making the hotel role-play time longer (starting with three minutes) since it was a more structured conversation where short-cuts were difficult to take, “wouldn’t it be just right to have 3 minutes for the hotel? The hotel you can’t skip parts because it follows a set manual.”

The model dialogue word count and time taken can be seen in the Table 4.10 for Study One, the hotel dialogue is clearly longer with 14 conversation turns, this compares to only 5-6 conversation turns in the café and directions dialogues.

Table 4.10 Model dialogue word count and speech rate for Study One

Unit	No. of words	Syllable count	Time in seconds	Conversation turns	Speech rate (syllables per minute)
Café	105	130	41	11 (6+5)	190
Directions	121	150	50	9 (5+4)	180
Hotel	190	236	87	28 (14+14)	163

Figure 4.10 shows the survey responses regarding the decreasing time role-play. 60% of participants found the decreasing time role-play useful for practising English, and 60% also found it difficult.

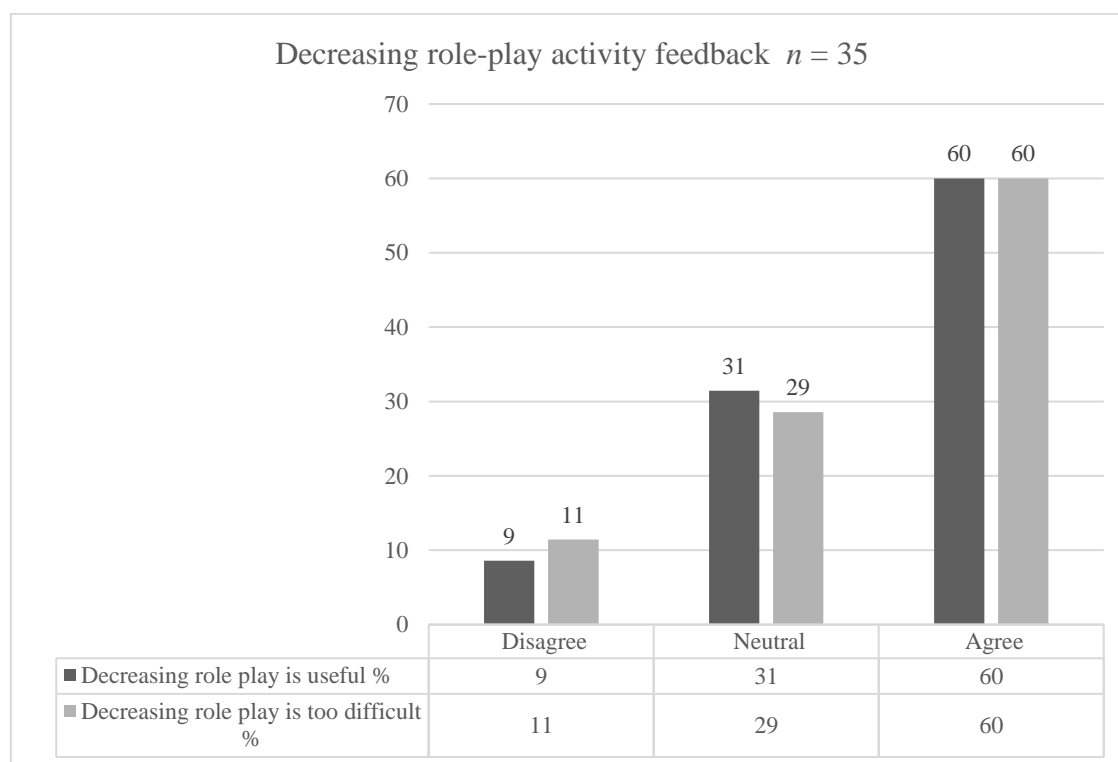


Figure 4.10 Activity feedback survey results for decreasing time role-play

My observations and participants' responses considered, it seemed that some adjustments to the decreasing time role-play would improve the effectiveness of the

activity. I decided to change the time allocations: starting with 90 seconds, decreasing to 75 seconds and 60 seconds for the restaurant and directions units, while starting the hotel scenario with three minutes, then decreasing to two and a half minutes and two minutes. I also decided to shorten the hotel dialogue to be more equal in time and words with the other two dialogues. The changes are described in detail in Chapter Five. The next activity to consider is the free role-play.

4.7.6 Free role-play

The free role-play required students to role-play a conversation with their partner for a scenario related to the model dialogue. The aim was that they would use the multi-word expressions in this new conversation. There was no model text given since it was supposed to be a self-generated conversation, but I did orally demonstrate an example of how they might use the multi-word expressions in a new scenario. I observed students struggling to start new conversations with their partners. Hana, Yumi and Yuki mentioned the lack of model dialogue meant they struggled with how to start the conversation, “free is the most difficult. First, I don’t know what is good to say.” Hana, Yumi and Yuki enjoyed the free role-play when they were partnered with friends or good speakers but found the free role-play quite difficult if they were partnered with someone who did not speak freely, “it depends who you are partnered with. It’s good if you are with someone who is into it, but difficult if you are with someone who doesn’t speak at all.” Hana said the free role-play was useful for English practice, while Yuki said, “it would be useful if I became able to do it.” Elaine said it was not such a useful activity because sometimes her partner did not understand her and vice versa, “sometimes my partner doesn’t understand me. I also don’t understand what that partner is saying, if it is like that we cannot communicate.” One comment from the activity feedback survey said that, “it was good practice to think and make my own sentences.”

Figure 4.11 shows that 61% of participants found the free role-play useful for practising English, but 46% also found it difficult.

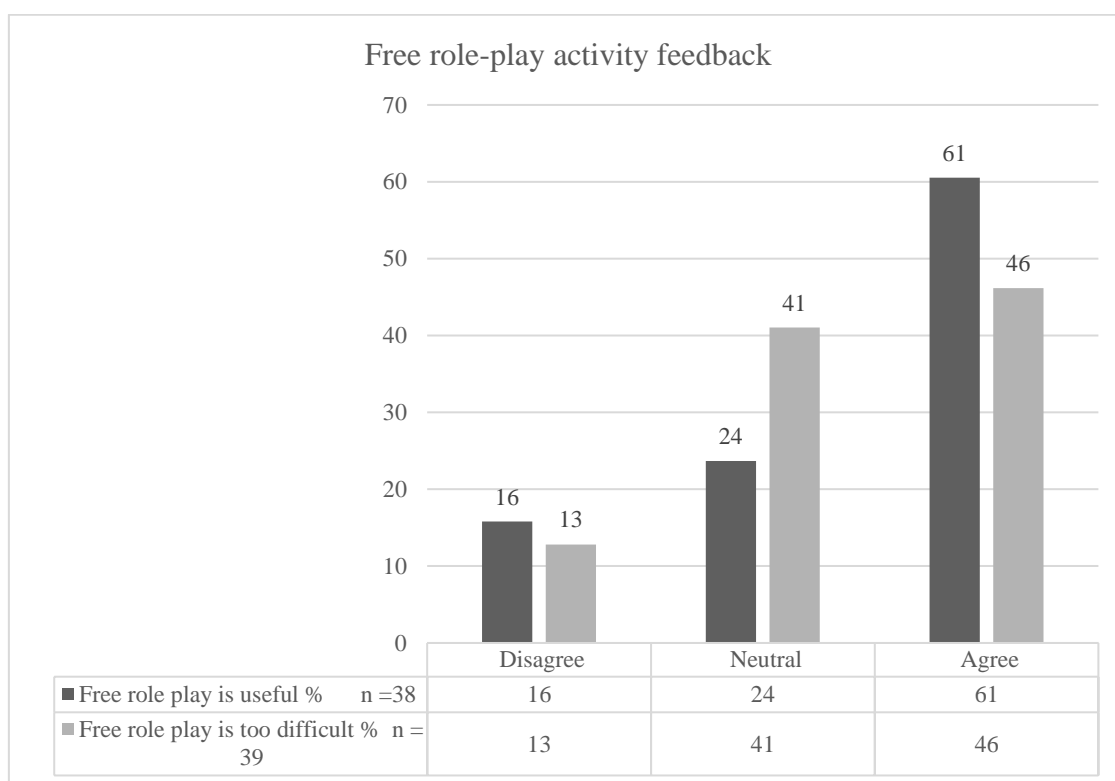


Figure 4.11 Activity feedback survey results for free role-play

It was clear from the observations and comments that the free role-play activity could be improved by a little more scene-setting to assist students in developing their free role-play conversation. Therefore, I decided to add more scaffolding and limit the choices for the free role-play in Study Two, description of these changes can be found in Chapter Five. The next activity for consideration is shadowing.

4.7.7 Shadowing

Shadowing was done using individual headsets so students could listen and repeat the words just after the audio, mimicking the pronunciation, rhythm and speed at their own pace. I had preconceived ideas that shadowing would not be valued in the classroom, because it was a simple mimicking activity with no creative production or communicative function. However, the focus group discussions reinforced the survey results indicating that shadowing was considered useful. Hana, Yumi and Yuki stated that shadowing made them read and speak faster and enabled them to learn pronunciation that they did not already know, “I think it is useful for speaking and reading faster.” Elaine stated that she thought shadowing was the most useful activity

because it allowed her to practise the rhythm and nuance of English, enhancing her ability to listen to and understand English, “It was the most useful. I could practise the feeling. If you can grasp the nuance and rhythm, you can listen to normal English.”

Figure 4.12 shows the activity feedback survey results for the shadowing activity, with 69% agreeing that shadowing was useful for practising English and only 13% finding it difficult.

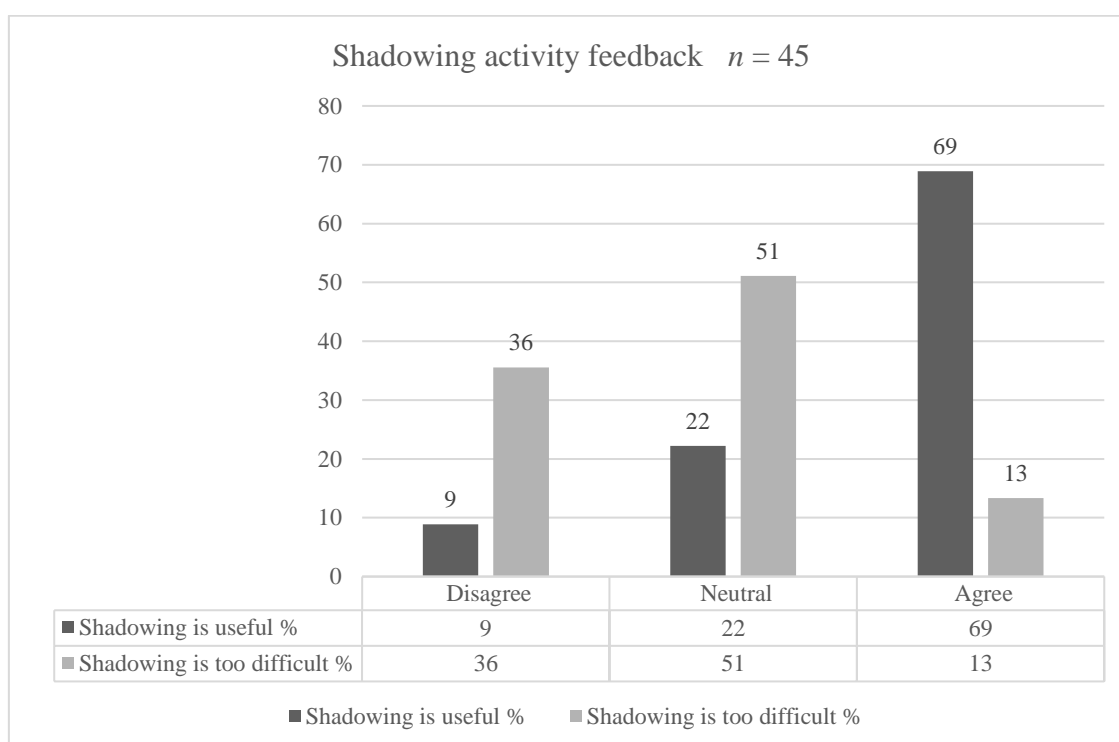


Figure 4.12 Activity feedback survey results for shadowing

Considering these observations and results, I decided that shadowing was appropriate, as it was, for the dialogue and EFL context. The next activity to consider is role-play.

4.7.8 Role-play train

My observation of the role-play was that it appeared to one of the more favoured classroom activities, because students were finally able to role-play the dialogue that they had listened to but not role-played up until this point. They could also read from their notes which gave them a bit of confidence. All four focus group participants

agreed that the role-play was useful since they were likely to find themselves in similar scenarios, “If you can remember and use it, it is useful.” When asked about the usefulness of the role-play, Yuki also noted that while she was mostly unable to complete the multi-word expressions cloze test the first time (pre-intervention test), she felt much more able the second time (post-intervention test) and thought this may have been due to the repetition of the role-plays, “I could do it so much better than the first time, so it was probably useful, I think it is naturally useful.” One comment in the activity feedback survey perhaps summed up the general response to this activity, “normally I don’t get the opportunity to converse in English, so I enjoyed it.” Figure 4.13 shows that 73% of participants found the role-play (using notes) useful, while 28% found it difficult. Therefore, I decided to continue to include the role-play in the Study Two without change. The final activity for consideration is phrase instruction.

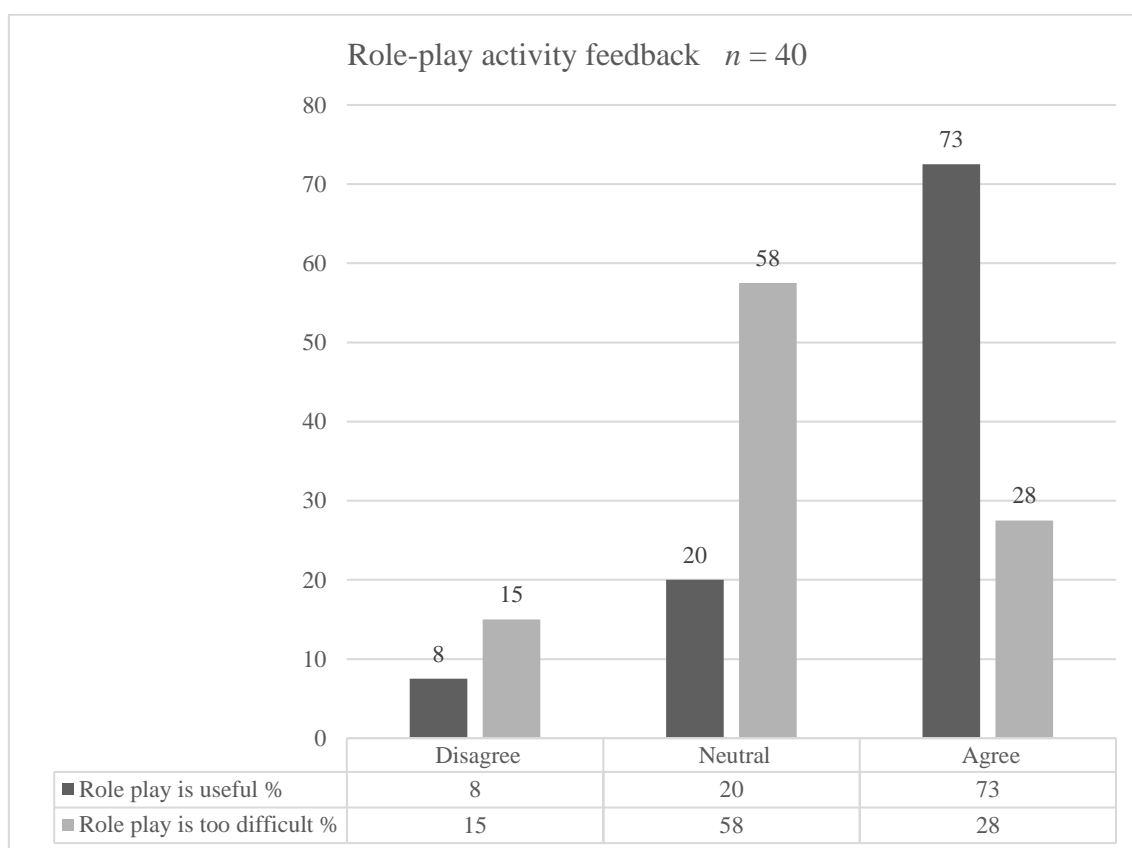


Figure 4.13 Activity feedback survey results for role-play

4.7.9 Phrase instruction

Phrase instruction was a receptive learning activity, it required only that participants look for and highlight the multi-word expressions in their transcript, practise pronunciation, and write the meaning or function of the multi-word expressions. As I introduced each of the multi-word expressions with their meanings, I noticed some students vigorously writing down the meanings on their transcript, while others did not. I wondered whether such explicit description of the meaning was necessary and whether students would understand the multi-word expressions without it. Hana, Yumi and Yuki noted that for some students phrase instruction would be revision rather than new learning, “I studied this kind of thing in high school, it is reminding me, feels like practice, it is useful for remembering.” Regarding phrase instruction, Elaine stated that, “if you know the phrase, you can make the sentence and you can change the subject etc. easily.” Figure 4.14 shows phrase instruction was considered useful by 91%, and only 9% considered it difficult.

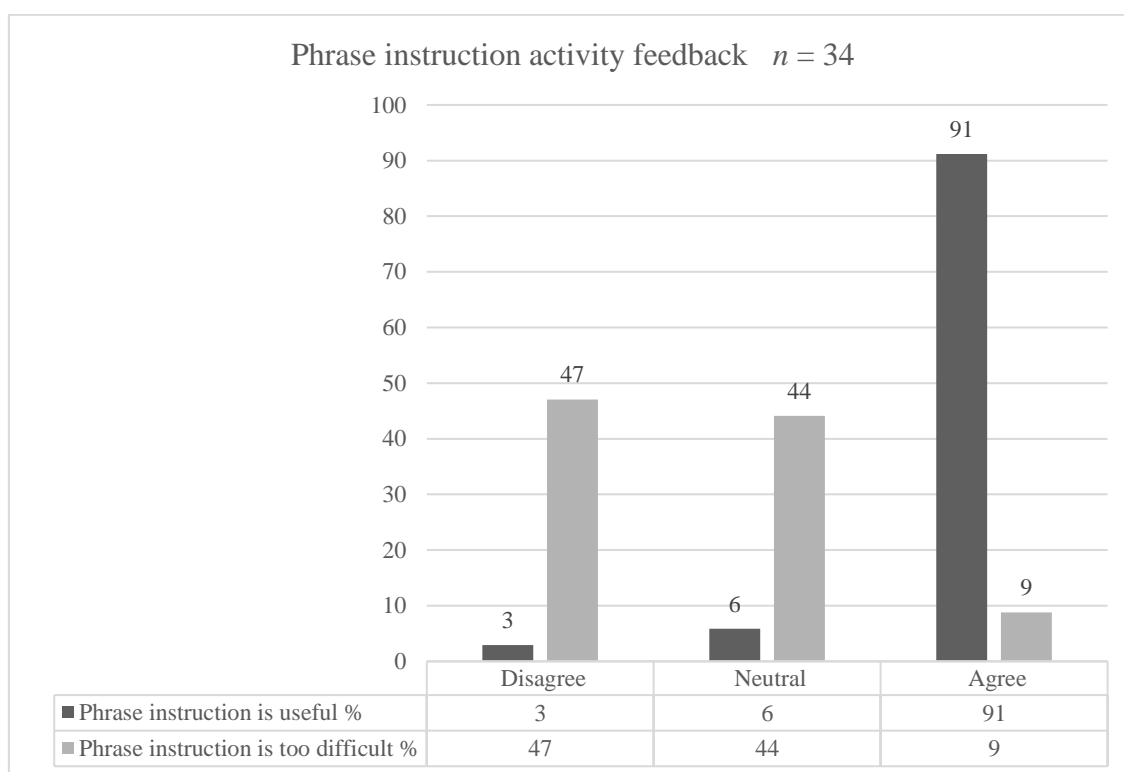


Figure 4.14 Activity feedback survey results for phrase instruction

One comment from the survey stated, “if I can memorise the phrase, I think I can use it, so it is good that it was taught.” It seems clear from the survey responses and the focus group feedback that phrase instruction was highly valued by learners. This activity therefore remained unchanged in Study Two.

Wood (2009) did not collect activity feedback from his participant. The narrative monologue model and Canadian ESL context from the 2009 study were different from this study, so it was to be expected that some activities that worked well in the classroom in Wood (2009) would not be as appropriate with the dialogue text and the EFL environment in Japan. Feedback from students from this study was used to inform improvements for the dialogue mode and EFL context in Japan. The feedback reported here may also be a useful reference for those interested in how certain classroom activities are perceived within a Japanese EFL teaching context.

4.7.10 Answering RQ. 6

Are adaptations necessary for the dialogue format and EFL context? If so, what? As detailed above, one activity (marking pauses) was found to be inappropriate for the dialogue mode, five activities required adjusting for the context (role-play recording, dictogloss, mingle jigsaw, decreasing time role-play, and free role-play), and three activities were considered appropriate as they were (shadowing, role-play, and phrase instruction). The adaption of classroom activities in response to this feedback will be further discussed in Chapter Five.

4.8 Study One results summary

Results presented here have addressed each research question in turn. The experimental intervention was found to be effective for building productive form and meaning knowledge of the 30 multi-word expressions. However, the intervention was less successful for increasing use of the expressions and speaking fluency. For discussion of what we have learnt through this replication see Chapter Seven. Feedback from participants and researcher observations revealed some areas where the experimental intervention activities could potentially be improved. Using the action research model, the researcher identifies an issue, acts to address the said issue, measures and reflects on the results, and makes changes to improve the method. Therefore, changes were made to the methodology for Study Two, the details of the changes and their impact on the study will be described in the next chapter, Chapter Five.

5 Chapter Five: Study Two method and results

In this chapter, I will introduce the adjustments I made to the overall methodology in response to limitations found in the initial study. Following an action research cycle approach (see Chapter 2.8 for more information on the action research concept), adjustments were made to 1) the data collection instruments, 2) the data collection process, and 3) the classroom activities. After describing the adjustments, I will introduce new research questions for Study Two. The results will then be reported along with participant feedback regarding the classroom activities.

5.1 Changes to improve data collection instruments

There were four changes made to data collection instruments in order to improve data collection and analysis. The reasons for each change (limitations found in the initial study) and a description of each adjustment follow. Table 5.1 summarises the changes made to improve the data collection instruments.

Table 5.1 Changes to improve data collection instruments

Changes made for Study Two
Expanded scenario for pre/post intervention dialogue recording
Activity feedback survey:
– Changed to query difficulty split by type (technical/linguistic/social)
– Changed to six-point Likert scale
– Added relevance and enjoyment feedback for each unit theme

5.1.1 Expand scenario for pre- and post-intervention dialogue recording

In Study One, my pre- and post-intervention dialogue recording instructions were for participants to role-play an interaction over a menu between friends at a café. However, after listening to the resulting dialogues, I realized that these role-play instructions were unlikely to prompt use of the full range of multi-word expressions (MWEs) that the experimental condition targeted. I considered it important therefore to expand the instructions for the spoken dialogue to include asking for and giving directions, and a service encounter. Therefore, I expanded pre- and post-intervention

dialogue instructions for Study Two, so that the first person was to ask the second person for directions to the nearest automated teller machine (ATM). The first person was then to ask the second person to go to lunch with them at the café. The two were then to imagine going to the café together and then to look at the provided menu. They were to order their food and take turns to play the role of the waiting staff. Due to the increased complexity of this scenario, I created bilingual (English and Japanese) instruction sheets and gave them out with the same menu as used in Study One (English only for the menu). Participants were given several minutes to read through the instructions with their partner before being told to record their conversation within a two to three-minute timeframe. The menus and instructions were collected again after the role-play had finished. The adapted role-play instructions are provided in Appendix 20.

5.1.2 Adapt activity feedback survey

In the initial study, the activity feedback surveys asked participants how difficult they perceived each experimental classroom activity to be. However, the focus group mentioned that ‘difficult’ could be experienced in various ways; such as linguistic difficulty (when writing notes for dictogloss), social difficulty (when approaching others in the mingle jigsaw), and also technical difficulty (when exporting to MP3 in the role-play recording activity). The initial survey question proved too broad and therefore unable to reveal information regarding the type of difficulty experienced. Also, the initial survey failed to probe the more positive aspect of whether the activity was enjoyable.

In order to discover more specifically the various difficulties encountered by participants, I split the survey statement about the difficulty of each activity into three statements, these were focused on 1) linguistic difficulty, 2) social difficulty and 3) technical difficulty. Since enjoyment is strongly associated with learning and achievement (Hagenauer & Hascher, 2014; Villavicencio & Bernardo, 2013), a separate question was also added to investigate each activity from a positive angle; to discover to what extent participants found the activity enjoyable. Therefore, participants were invited to rate each experimental class activity on perceived usefulness for practicing

English, difficulty (linguistic, social and technical) and enjoyment. See Figure 5.1 for the amended activity feedback survey questions. In the next section, I will explain why and how I adjusted the Likert scale.

Please indicate to what extent the following statements are true

	Strongly disagree	Disagree	Somewhat disagree	Somewhat agree	Agree	Strongly agree
Mingle jigsaw was socially difficult	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mingle jigsaw was technically difficult	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mingle jigsaw was linguistically difficult	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mingle jigsaw was useful for practicing English	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mingle jigsaw was enjoyable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Comment

Figure 5.1 Adjusted activity feedback survey example

5.1.3 Change from five-point Likert scale to six-point Likert scale

The five-point Likert scale used in the activity feedback survey in Study One allowed a neutral response, and most participants tended to choose either the neutral (three) or positive (four) response on the scale. Since choosing three indicated a neutral position, it was not particularly informative. It is possible that participants chose responses that allowed them to avoid expressing extreme opinions, a cultural tendency for Japanese (Lee et al., 2002; Wang et al., 2008). Therefore, in order to allow participants to express an opinion without feeling extreme, a six-point Likert scale was used in the follow up study, forcing a choice between agree or disagree. The new six-point Likert scale can also be seen in Figure 5.1. In the next section, I will explain why and how I added a question to the activity feedback survey regarding the unit themes.

5.1.4 Add feedback for each thematic unit

I wanted to confirm whether the participants felt the themes in the experimental condition were relevant for their future anticipated use of English, and whether they

enjoyed the unit themes. However, there was no survey question asking participants about unit theme relevance or whether the unit themes were enjoyable. Therefore, in order to check the relevance and enjoyment of the unit themes, two questions were added to the final activity feedback survey; 1) to ask learners how useful they perceived each thematic unit to be for their future, and 2) asking whether the unit was enjoyable. See Appendix 21 for an example screen shot of the adjusted survey format with English translation provided.

5.2 Changes to improve data collection methods

As the researcher, I also noticed limitations with the data collection methods. I therefore decided to make five adjustments to try to improve the data collection methods for Study Two. A summary of the changes made to data collection methods can be seen in Table 5.2. These changes and their anticipated impact to the study are described in more detail in this section. At the end of this section I describe how I used PRAAT software (Boersma & Weenik, 2017) to analyse the spoken dialogue recordings in Study Two.

Table 5.2 Summary of changes to data collection methods in Study Two

1 Familiarise all groups with dialogue recording task before pre-test
2 Individual speakers self-identify at start of recordings
3 Instruct participants to record pre-intervention scores for LVLT test and MWE cloze test
4 All participants given forms to indicate their willingness to participate in focus group
5 Have participants export audio file as a WAV file

5.2.1 Familiarise all participants with dialogue recording task

In Study One, the experimental group had been given one experience of recording conversation during class before the pre-measure recording, whereas the control group had not. This may have caused an imbalance in the two groups’ gains from pre- to post-intervention fluency measures as a result of task familiarity. Therefore, while there was

no control group in Study Two, I made note to ensure that both the control and experimental groups would have experience with recording before starting Study Three.

5.2.2 Individual speakers self-identify at start of recordings

In Study One, speakers did not self-identify in the pre- and post-intervention dialogue recordings. The lack of identifying information on the audio made it very difficult to determine who was speaking, which meant that I had to do analysis by dyad rather than individual, reducing the sample size by half. Therefore the risk of declaring no effect from the experiment when there actually was (Type II error) and/or having sampling error was increased unnecessarily (Field, 2009, pp. 68–69). Therefore, in Study Two I added an instruction for participants to state their names, student numbers and role at the beginning of the audio recording. I hoped that this would make it easier to associate utterances with individual speakers enabling analysis to be done for each individual, thereby making the most of the potential sample size.

5.2.3 Instruct participants to record pre-intervention scores

In Study One, participants understood the research test scores were unrelated to any class grading. Post-intervention scores for the LVLT test and multi-word expressions cloze test likely held little meaning to participants as they had no way to make a comparison with their pre-intervention scores. As a result some participants appeared to disengage with the tests. Participants had momentarily seen their pre-test LVLT and multi-expression cloze scores six classes earlier, but it was unlikely that they remembered them. Therefore, in order to boost engagement with the tests in Study Two, participants were instructed to write each vocabulary listening level score and their multi-word expression cloze score on the inside of their name card. Before doing the post-tests, they were encouraged to check their recorded pre-test scores so that they were conscious of the comparison they could make. I hoped that raising self-awareness in this way would motivate participants to engage fully with the post-tests.

5.2.4 Change in focus group recruitment method

In the first study, I asked those who were interested in participating in a focus group to see me at the end of class to receive information and forms. The response was

disappointing, and I thought perhaps this method required too much extra effort from students. Therefore, in an effort to make it easier for students to indicate their willingness to participate in a focus group and thereby increase participation, I gave the information sheet along with a consent form to all class members in the last five minutes of class in the 5th week. I asked them to read the information sheet about the focus group and indicate their willingness or not to participate on the consent form and put the completed in a box at the front of the room as they left.

5.2.5 Dialogue recording: From MP3 to WAV format

In Study One, dialogues were audio recorded and exported to MP3 file format because MP3 files took up less data storage space than WAV files. However, the MP3 format was not compatible to use with PRAAT (Boersma & Weenik, 2017). Using PRAAT would have enabled more precise measurements of fluency. For example, pausing phenomena could have been measured, which would have allowed phonation time ratio to be computed. Therefore, in Study Two, audio recordings were exported and saved as WAV files in order to be compatible with PRAAT analytical software. WAV files take up more storage space but are more detailed. The next section explains how I used PRAAT for analysis of the dialogue recordings in Study Two.

5.2.6 PRAAT analysis details

Since PRAAT (Boersma & Weenik, 2017) was not able to be used in Study One, the specifics of how it was used will be described here. The WAV files were played and visualized using PRAAT TEXTGRID. Speaking and pausing segments were identified and labelled on a single sound tier to identify each speaker's speaking time and pause time. Unlike the first study, the phonation time ratio without pauses was thus able to be calculated, giving a more precise measurement of speech phenomena.

Speakers introduced themselves by name, student number and role at the beginning of the recording so that they could be identified for analysis. These introductions and the pause that followed were not transcribed or included in speaking time. They were labelled as *intro* on TEXTGRID. Speaking time started when a speaker initiated the conversation with a filler or words after the introductions had finished.

Speaking time for a speaker was labelled in PRAAT with an initial or the first letters from their name. For example, if a participant was called Tomoko, then her speech would be labelled with *t*. Filled pauses (for example *um*, *ah*, *er*) and repetitions were included as speaking time. A pause was defined as silent non-speaking time greater than 0.25 seconds and labelled as *p* (following N. H. de Jong et al., 2012; Tavakoli et al., 2016). Pauses between speaking turns were split evenly between the speakers, while pauses within a speaking turn were counted as pauses only for that speaker. For example, a pause between Tomoko's words would be labelled *tp* to identify it as one of her within turn pauses. When the conversation stopped for long periods or Japanese language was spoken, the segment was labelled as *tran*; short for transition time. Speaking time ended when the speakers stopped performing the role-play, and this was labelled *end*.

The process of segmenting and labelling speaking and pause time on PRAAT down to 0.25 seconds took a considerable amount of time, a three-minute conversation could take over one hour for speech marking/time-stamping in PRAAT. In order to speed up the speech segmentation process in PRAAT, I applied for and was approved ethics permission to hire technical support to listen to the audios and make PRAAT TEXTGRID files showing the speaking segmentation. I used the online freelancing site UPWORK (upwork.com) to recruit someone with skills and experience using PRAAT to do preliminary speech segmentation. The chosen person signed a confidentiality agreement before commencing the work, the agreement can be seen in Appendix 22. The large audio WAV files could be shared privately with the technical support person using an online file sharing platform called Dropbox. I gave the technical support person an example audio file with an annotated PRAAT Textgrid file so that she could see how to annotate (label) and segment the speech. She was then given two files to segment (time-stamp) and label as instructed, I checked the segmentation and gave feedback on how to improve the labelling. I then sent her further files to segment and label.

For Study Two, there were eight audio files to segment (four pre-measure and four post-measure). After the files had been segmented and annotated, I further edited the files to ensure that no phenomena of 0.25 seconds or greater were left out and that

the segments were marked correctly for speaker and content. While editing, I had PRAAT and TextPad files open simultaneously so I could check the segmentation against the transcription and vice versa. After I had finished editing all files, I checked the complete set of files a third time to ensure consistency in the labelling between the files segmented and transcribed earlier and those done later as several months had elapsed from the beginning to the end of this process. I found that the first half of the files were not as finely segmented as the second half, so I made edits to bring all files up to the same labelling precision. In particular, in instances where both speakers were talking simultaneously, I split the speaking time between the speakers. Drawn out word endings were included as speaking time and did not break a run. I included meaningful repetitions in the trimmed transcripts, while pure repetition (within the speaker's turn) was not included in the trimmed transcripts.

Once the PRAAT TEXTGRID files had been labelled, I was able to run a script (created by Lennes, 2002) in PRAAT to automatically calculate the total duration of the labelled segments. In this way, I could measure individual speaking time, pause time and other labelled segments.

5.3 Changes to improve classroom activities

As well as making changes to the data collection instruments and methods, I also made changes to the classroom activities. As the classroom teacher and the researcher, I considered the effectiveness of the classroom activities based on my own observations and the feedback from participant surveys and participant focus groups. Adjustments were made to improve the classroom learning experience as detailed in Table 5.3, and further explanation follows.

Table 5.3 Outline of changes to classroom activities for Study Two

1	Remove marking pauses activity
2	Provide structure and class time for reflection on recorded role-play
3	Change dictogloss to group work and add more scaffolding with 5-minute time limit
4	Add 5-minute time limit to mingle jigsaw
5	Shorten hotel dialogue
6	Display MWEs for decreasing time role-plays and free role-play
7	Shorten decreasing time role-play limit to 1.4 – 1.2 – 1 (previously 2 – 1.5 – 1)
8	Remove choice of scenario for free role-play, add planning time and more scaffolding

5.3.1 Remove marking pauses activity

In Study One, both focus groups said that the marking pauses activity had little meaning since they could not hear any pauses. They therefore resorted to mark meaning boundaries instead. The marking pauses activity is perhaps better suited to a monologue narrative model, such as in Wood (2009). The dialogic speech models in this study had few clear pause phenomena since the conversation turns were short. Considering this feedback from the focus groups and my own observations, I removed the marking pauses activity for Study Two.

5.3.2 Provide structure and class time for reflection on role-play recording

In Study One, participants were encouraged and able to listen back to their role-play recordings. However, only one out of the four focus group participants actually reported listening and reflecting on her recording. There is little point in recording if you do not then listen back and reflect on your production. It would appear that leaving reflection up to students in their own time meant that the majority did not reflect on the recording. Survey and focus group feedback suggested that recording was a lot of effort and technically complicated. Listening back to the recording would have added validity to the effort and time invested in the actual recording. Therefore, in the follow-up study,

the time saved by removing the marking pauses activity was re-allocated to listening back to the role-play recording and self-reflecting on pausing behaviour and speech clarity (pronunciation) in class. The reflection was structured with the use of a bilingual fluency reflection worksheet that was collected by the teacher and re-distributed with every recording in order to raise learners' self-awareness of fluency development. See Appendix 23 for a copy of the fluency reflection worksheet.

5.3.3 Add scaffolding to dictogloss hand-out

In Study One, participants rated the dictogloss activity as one of the more difficult activities. Participants in the focus groups suggested pausing the audio periodically to allow time for writing, or to provide more of the text on the handout. Extra scaffolding to make this activity more achievable was obviously needed. Therefore, I edited the dictogloss hand-out to provide more guidance for writing in the form of clear boxes for each conversation turn. A section to take notes was also provided above the conversation turn boxes. See Appendix 24 for an example of the adjusted dictogloss hand-out from the café unit. Also, pair work was changed to group work as follows: students listened to the audio four times; on the first playback I instructed them just to listen and not to write anything, in order to focus on the overall meaning. On the second, third and fourth playbacks, they were instructed to write the words they could hear. After the fourth playback, students were told to compare notes in small groups of four and work together to reconstruct the text. The group interaction was limited to five minutes, with a class timer on display. These changes were made to encourage faster group interaction, motivation and engagement with the task.

5.3.4 Set five-minute time limit for mingle jigsaw

In Chapter Four, it was clear from focus group feedback regarding the mingle jigsaw that hesitation and doubt about approaching some class members was a hindrance to actively collecting the expressions and completing the activity. Mingle jigsaw was not a time pressured activity in Study One, neither was it in Wood (2009) (Wood, personal communication, 14 March 2018). However, in order to re-focus students from dithering about who to ask and instead concentrate on collecting the

expressions, I decided to apply time pressure by displaying a five-minute countdown timer during the activity in Study Two.

5.3.5 Shorten hotel model dialogue

I observed that the time limits set for the decreasing time role-play were too long for the café and directions themed role-plays, but too short for the hotel themed role-play. When I compared the timing of the model dialogues, it was clear that the hotel dialogue was longer. Therefore, I shortened the hotel model dialogue to be more comparable with the other two models for Study Two. See Table 5.4 for a comparison of the word counts and timing, note that Hotel (shortened) while still being longer is closer in length to café and directions than Hotel (original). The shortened hotel model dialogue along with vocab profiling can be seen with the other model dialogues in Appendix 9.

Table 5.4 Word count, time and conversation turns for model audio tracks

Unit theme	Word count	Time	Conversation turns
Café	105	0:41	11 (6+5)
Directions	121	0:50	9 (5+4)
Hotel (original)	185	1:27	28 (14+14)
Hotel (shortened)	141	1:12	14 (7+7)

5.3.6 Display target multi-word expressions

One of the aims for the decreasing time role-play and free role-play was that learners would use the target multi-word expressions in their conversations. However, when participants were not allowed to refer to their notes, I observed them reverting to the safest way they knew to communicate, which often did not include using the freshly learnt multi-word expressions. In order to encourage greater use and therefore learning of the multi-word expressions, especially in the case where students are unable to retrieve the expression from memory, it would help to have the target expressions visible. Display of target expressions on the classroom screens during these activities

would serve to increase the number of encounters learners have with the expressions. Frequency of encounter is known to be an important factor in vocabulary learning (Joe, 1998). Therefore, in Study Two the target expressions were displayed on the class screens during the decreasing time role-play and free role-play. Participants were not allowed to look at their notes, but they were able to see the target expressions on the screen if they wanted to.

5.3.7 Shorten initial timings for decreasing time role-plays

The time allowed for the decreasing time role-plays was found to be too long according to focus group feedback. The time given was actually serving to slow speaking rather than speed it up, as speakers tried to fill the time by speaking slower. The decreasing time role-play started with 120 seconds (2 minutes), decreased to 90 seconds (1 minute 30 seconds) and finally to 60 seconds (1 minute). In particular the first- and second-time allocations were a bit too long. Therefore, in Study Two I reduced the time available for the first two iterations. The decreasing time role-play therefore started with 100 seconds (1 minute 40 seconds), decreased to 80 seconds (1 minute 20 seconds) and finally to 60 seconds (1 minute).

5.3.8 Reduce choice in free role-play

The free role-play instructions presented participants with a choice of two or three related scenarios to role-play. I observed that students were often paralyzed by the choice and that they also appeared to need more guidance to get started with a role-play. Feedback from the focus groups regarding the free role-play suggested that more direction was necessary to start these conversations. Therefore, I removed the choices of scenario, and gave pictures and scaffolding for a single scenario that they were all to role-play. For example, the free role-play for the café unit became: You are at a convenience store, discuss what you want to buy. I also added one-minute planning time, where students were allowed to speak in the L1 with their partner to decide roles and the main ideas of the conversation.

I have described changes made to the data collection instruments, methods and classroom activities based on feedback from Study One. These adjustments were made

as part of the action research framework and were intended to improve the quality of the research and classroom learning experience. Now the research questions for Study Two will be introduced, followed by information on a new cohort of participants and the results for Study Two.

5.4 Study Two

This section introduces the research questions, participants and results for Study Two. Each results section addresses a research question, starting with quantitative learning results and finishing with qualitative feedback from participants regarding the classroom learning experience. The research questions will now be introduced.

5.4.1 Research questions

There were five research questions that I wanted to investigate for Study Two. These research questions were:

- 1) Is there an increase in multi-word expression knowledge in Study Two that supports the results from Study One?
- 2) Does multi-word expression use in Study Two increase more than Study One?
- 3) Does spoken fluency increase in Study Two more so than in Study One?
- 4) Does non-targeted vocabulary knowledge remain unchanged in Study Two as it did in Study One?
- 5) Which activities are considered the most useful by learners?

5.4.2 Study Two participants

Participants were again recruited from intact EFL classes in the same engineering university as in Study One. Due to scheduling limitations, there was only an experimental group (no control group) in this study. Only data from those who consented to participate are included in the analysis and reporting. The same language learning background survey was given as in Study One, though a question was added to ascertain the gender of participants. As with the original group, the second group of participants were also predominantly male. Of 22 participants, only 2 were female. All participants were Japanese university students of 18-19 years old. Table 5.5 compares participants' TOEIC score range (scores were self-reported as a range rather than as

exact scores), with the score range of the initial experimental participants. Since normality was doubtful in both data sets of TOEIC scores, a Mann-Whitney test was used to check the comparability of this second experimental group with the first experimental group. Based on the numbers of students in each score range there was no significant difference found in TOEIC score range (proficiency) between the first group ($n = 52$) $Mdn = 2.00$ and the second group ($n = 22$) $Mdn = 2.00$, $U = 505.0$, $z = -.865$, $asympt\ p = .387$, $r = .10$. The TOEIC score ranges correlated positively with the Listening Vocabulary Levels Test (LVLT) scores before the intervention: $rs(22) = .438$, $p = .041$. The LVLT score therefore could be considered as an indicator of language proficiency.

Table 5.5 Participant numbers by group with TOEIC score percentages

TOEIC score (IELTS equivalent)	Experimental 1 ($n = 52$)	Experimental 2 ($n = 22$)
(1) 10-250 (IELTS 0-1.5)	13.5%	31.8%
(2) 255-400 (IELTS 2-3.5)	57.7%	36.4%
(3) 405-600 (IELTS 4-5)	25%	31.8%
(4) 605-780 (IELTS 5.5-7)	3.8%	0%

5.4.3 Measuring effect

Participants were pre-tested before participating in experimental classroom activities for six weeks, and then tested again after the intervention. The method from Study One was adapted as described in the adjustment sections above and the results are reported below. Data distributions were investigated and are only mentioned when they deviate from normal. The alpha significance level was set at 0.5 and effect sizes were also calculated. Table 5.6 shows the various measures that were used and their purpose. The following sections 5.5 - 5.9 present the results.

Table 5.6 Measures with purpose

Measure	Purpose
Multi-word expression cloze scores	Measure productive form and meaning knowledge of multi-word expression
Use of multi-word expressions in conversation	Measure ability to use multi-word expressions
Speech rate	Measure fluency
Phonation time ratio	
Articulation rate	
Mean length of runs	
Listening Vocabulary Levels Test (LVLT)	Measure general unrelated aural vocabulary knowledge

5.5 Multi-word expression cloze-test gain scores

The first research question asked whether there was an increase in multi-word expression knowledge in Study Two that would support the results from Study One. In order to investigate this question a comparison was made between the gain scores on the multi-word expression cloze test in Study One and Two.

5.5.1 Multi-word expression cloze-test gain scores compared with Study One

Gain scores for multi-word expression knowledge were calculated by subtracting the multi-word expression cloze pre-test score from the post-test score. For Study Two, a significant difference with a large effect size was seen between the pre- ($M = 8.47$, $SD = 2.58$) and post-test results ($M = 16.10$, $SD = 5.10$), $t(21) = -8.39$, 95% CI [-9.52, -5.74], $p < .001$, $d = 1.89$. As stated in Chapter Four, the highest possible score for this test was 30. When compared, there was no statistical difference between Study One experimental group gain scores $n = 45$ ($M = 7.65$, $SD = 4.08$) and Study Two $n = 22$ ($M = 7.63$, $SD = 4.27$), $t(65) = .013$, 95% CI [-2.14, 2.17], $p > .05$, $d = .005$. This result suggests that the changes made to the classroom activities did not significantly influence the effectiveness of the activities for building productive knowledge of target

multi-word expressions. Therefore, in response to the first research question, this result supports the findings from Study One, that the fluency workshop intervention builds productive knowledge of form-meaning links of multi-word expressions. See Chapter Seven for further discussion.

5.6 Use of multi-word expressions in conversation

The second research question regards the use of multi-word expressions in conversation. This section reports comparisons between Study One and Two multi-word expression use by dyad. The range of multi-word expressions used is compared between the two studies and finally individual use of multi-word expressions for Study Two is compared before and after the intervention. A list of the multi-word expressions used by participants in the pre- and post-intervention dialogues can be found in Appendix 25. The measurement of multi-word expression use was extended to include combinations of multi-word expressions such as *do you like*, which was counted as an extension in Study One. Extended multi-word expressions were not counted in Study Two. The counting of multi-word expressions in Study Two encompassed more variation than the Study One MWE count but was not as encompassing as the Study One EMWE count. We naturally vary our use of multi-word expressions, and as such the slightly less restricted counting method in Study Two is more sensitive to reveal learning and ability to use the multi-word expressions targeted in the class activities. This difference in measure presents a limitation when making comparisons between Study One and the following studies and is listed among limitations in Chapter Eight.

5.6.1 Speech analysis of eight participants

Conversations were transcribed only for those participants who had the same partners for pre- and post-measure recorded dialogue, and only those who played the same roles in both recordings. Using these criteria, there were only four pairs (eight individuals) able to be included in the speech analysis. The small sample size should be kept in mind when interpreting the results. Variances in individual data in such a small data set are likely to have a large impact on the group average. Unless otherwise stated, the data is based on untrimmed transcripts treating the transition times as pause time.

5.6.2 Comparing multi-word expression use with the first study

In the first study, the use of multi-word expressions was investigated in the trimmed conversations of only 15 dyads. Combining the speakers into four dyads in Study Two, a comparison was made between the trimmed multi-word expression use gain scores between the Study One experimental group ($Mdn = 4.0$) and Study Two experimental group ($Mdn = 5.5$). Since the first study group had an outlier which distorted the distribution with a significant Kolmogorov-Smirnov score $D(15) = .220$, $p = .050$, a Mann Whitney test was used to compare the groups. There was no significant difference found between the first and second experimental group for multi-word expression use gain scores $U = 29.5$, asymp $p = .96$, $r = .01$. In Study One, there was a significant difference found between use of multi-word expressions in the pre- and the post-measure ($d = .81$). In Study Two, the dyads averaged 10 words from multi-word expressions ($SD = 4.97$) in the pre-measure, while the average increased to 14.00 in the post-measure ($SD = 3.37$). The effect size was large, suggesting that this change is significant. In this case the effect size, which is not influenced by sample size, is the most appropriate outcome to notice $t(3) = -1.96$, 95% CI $[-10.5, 2.5]$, $p = .15$, $d = .94$. Both Study One and Study Two have large effect sizes when comparing use of multi-word expressions pre- to post-measure. Therefore, Study Two results appear to agree with those of the first study when considered by dyad, that use of multi-word expressions increases for participants after the multi-word expressions focused activities. These results will be further discussed in Chapter Seven.

5.6.3 Comparing range of multi-word expression use with the first study

It was expected that the expansion of the dialogue instructions in Study Two would allow use of a wider range of multi-word expressions. In the previous study there were 12 multi-word expressions that were partially or fully used by participants in the post-measure dialogue. In contrast, there were only nine multi-word expressions that were partially or fully used by participants in the follow-up study. It is not clear therefore, whether the decision to expand the role-play scenario encouraged the use of a wider range of multi-word expressions. In fact, expanding the scenario may have served

to reduce the overall speech rate as participants had to process more complex instructions regarding their roles.

5.6.4 Individual use of target multi-word expressions in conversation

The comparison above was based on the trimmed speech analysis of dyads, as analysis by dyad was a limitation of Study One. However, in Study Two, I was also able to analyse speech by individual because speakers self-identified at the beginning of the audio recordings. This analysis by individual revealed more precise information regarding change in speaking for each participant. Based on individual analysis, there was a 38.1% increase in use of target multi-word expressions from pre- ($M = 5.25$, $SD = 4.62$) to post-measure untrimmed conversation ($M = 7.25$ $SD = 3.96$). A paired samples t -test, however, showed that this increase was not statistically significant with a small to medium effect size: $t(7) = -1.79$, 95% CI $[-4.64, .644]$, $p = .117$, $d = .46$. With such a small sample size the lack of significance is not entirely surprising. As a point of comparison, based on item analysis from the multi-word expression cloze test, there was an average increase of 23% from pre- to post-test multi-word expression knowledge. Table 5.7 shows the change in use of target multi-word expressions from pre- to post-measurement by individual, measured by words used in multi-word expressions (MWEs). Speakers are labelled in relation to their speaking partners, therefore A1 and A2 were speaking partners. Use of target multi-word expressions was counted from words used in target multi-word expressions exclusively. Six speakers (C1, B1, A1, A2, B2, and C2) used more words from target multi-word expressions in the post measure, while there was no change for one speaker (D2) and a decrease in use for one (D1).

Table 5.7 Change in number of words spoken as part of multi-word expressions

ID	Pre MWE use	Post MWE use	Pre to Post Change
C1	4	10	+ 6
B1	6	11	+ 5
A1	6	10	+ 4
B2	2	4	+ 2
C2	0	2	+ 2
A2	7	8	+ 1
D2	2	2	0
D1	15	11	-4

(counted from untrimmed words used in MWEs)

Research question two asked whether use of target multi-word expressions in Study Two increased more so than in Study One. Since there was no statistically discernible difference in gain scores between the two experimental groups, the answer is that Study Two did not result in increased use of target multi-word expressions any more so than Study One.

5.7 Spoken fluency

Research question three asked whether spoken fluency increased after the intervention, more so than in Study One. In response to this question measures of fluency from Study Two are considered in the following order: speech rate, phonation time ratio, articulation rate and mean length of runs. Where possible, comparison with Study One is made.

5.7.1 Speech rate compared with Study One experimental group

The speech rate is calculated as the sum of syllables spoken divided by seconds (speaking time including pauses) x 60 (Tavakoli, 2016; Towell et al., 1996). Therefore,

the speech rate represents the number of syllables spoken per minute. In the original study, there were 15 experimental dyads, while there were only four in Study Two. In the first study, the untrimmed all-inclusive speech rate increased significantly from pre- ($M = 50.82$, $SD = 18.13$) to post-measure ($M = 63.29$, $SD = 18.54$), $t(14) = -3.64$, 95% CI $[-19.8, -5.1]$, $p = .003$, $d = .68$ (though the difference with the control group was not significant). In the follow-up study there was a similar effect size for pre- ($M = 59.45$, $SD = 17.14$) to post-measure changes ($M = 72.85$, $SD = 23.96$), $t(3) = -.686$, 95% CI $[-75.6, 48.8]$, $p = .542$, $d = .64$. The non-significant p is hardly surprising considering that there were only four dyads in Study Two and p values are affected by sample size (Larson-Hall, 2015, p. 143). An independent t -test where equal variances were not assumed showed no significant difference between the original experimental group ($M = 12.47$, $SD = 13.28$) and the second experimental group ($M = 13.40$, $SD = 39.08$) in terms of gain scores from pre- to post-measure for speech rate $t(17) = -.047$, 95% CI $[-62.0, 60.2]$, $p = .966$, $d = .03$. Therefore, the follow-up study results appear to support the Study One results regarding speech rate.

5.7.2 Individual speech rate change for Study Two

The speech rate change for individual speakers in Study Two is shown in Table 5.8 along with their self-reported TOEIC proficiency score range. If a speaker had long pauses, their speech rate would be lower. When speech rate was investigated through individuals (as opposed to the dyad analysis given above), the average speech rate ($M = 42.99$, $SD = 13.35$) increased by 10.6% in the post-measure ($M = 47.53$, $SD = 19.15$). However, a paired samples t -test showed that this increase was not statistically significant $t(7) = -.641$, 95% CI $[-21.3, 12.2]$, $p = .542$, $d = .28$.

Table 5.8 Speech rate before and after treatment

ID	TOEIC score range	Pre speech rate	Post speech rate	Change
D1	10-250	43.58	80.69	37.11
D2	10-250	21.86	36.81	14.95
C2	10-250	25.96	40.73	14.77
B1	405-600	56.64	67.72	11.07
C1	255-400	47.46	57.34	9.88
A1	255-400	49.56	37.79	-11.77
B2	405-600	39.44	23.96	-15.49
A2	405-600	59.09	35.23	-23.85

(untrimmed all-inclusive)

I was expecting the untrimmed all-inclusive speech rate to increase from pre- to post-measure. Indeed, the speech rate did increase for five out of the eight participants. There was, however, a decrease in speech rate for three participants. Many factors could cause a decrease in speech rate, for example distractions from the task such as overhearing other conversations, personal interaction conditions, and trying to retrieve words or phrases that had been learnt. The conversation between A1 and A2 was particularly slow and drawn out. In the post-measure conversation A2 appeared distracted as he took a long time to respond to his partner. A2 had the largest decrease in speech rate of the participants. This decrease in speech rate may have been due to trying to retrieve expressions that he thought he should use. It is interesting that those who improved their speech rate the most were those with lower TOEIC score ranges, perhaps indicating that they had more potential for improvement than those with higher proficiency.

Interestingly, there were only two instances where participants used an exact target expression in this second study. Extract 5.1 shows the transcript for speaker A2,

who provided one of those instances; he managed to use the exact full four-word target expression in his post-conversation: 'I think I'll (have).' This utterance was preceded by a filled pause, hesitations and repetition.

Extract 1 Speaker A2

ah

(pause 2.5 seconds)

I

(pause 2.3 seconds)

I think I'll have

(pause 1.4 seconds)

chicken

A2 did not use or show any knowledge of this expression (or the related expression 'I will have the') in the pre-test (both spoken dialogue and cloze-test), but he chose to use this expression in his post-measure dialogue and was able to type the words correctly in the post-cloze test under time pressure. In this instance, it appears that the experimental class activities resulted in both knowledge of form and productive ability to use this expression. However, in this case, use of a multi-word expression was associated with a fall in speech rate. The fall in speech rate indicates that while procedural knowledge (knowledge of how to use the expression) was developing, it was not yet automatized to the point that it would improve fluency.

In contrast to A2 whose speech rate decreased the most, D1 had the largest increase in speech rate of the eight participants. Extract 5.2 shows the transcript from speaker D1, who was the other participant who used an exact target expression in the post-measure 'can you tell me'. Unlike A2, he had no hesitation before using the target expression.

Extract 2 Speaker D1

how I get

(pause 0.4)

to

(pause 0.4)

ATM can you tell me

D1 had only been able to type the word ‘you’ as part of the multi-word expression *can you tell me* correctly in the pre-intervention cloze test whereas he typed the whole multi-word expression correctly in the post-intervention cloze test. He also showed ability to partially use the expression in the pre-measure conversation by saying “can you help me,” D1 showed improvement in his knowledge of form and productive ability with the target multi-word expressions as a result of the experimental classroom activities, and this increase in multi-word expression knowledge and use was associated with an increase in fluency.

It is conceivable that the ability to improve fluency might be related to proficiency. From the background data collected on A2 and D1, it appears that A2 actually had a higher proficiency to start with, claiming a TOEIC score range 406-600, 4-5 years of private lessons, and having spent between 15-30 days overseas. His pre-test vocabulary levels score was 54 but dropped to 53 in the post-test. A2 scored 8.5 on the pre- MWE cloze test and 17.08 on the post-cloze test. In contrast, D1 claimed a TOEIC score range of 10-250, with no private lessons or time overseas. His pre-test vocabulary levels score was 45 which increased to 50 in the post-test. His MWE cloze test score moved from 7.83 to 16.75. So, it appears that D1 was starting at a lower proficiency level, perhaps allowing him more room for improvement.

5.7.3 Phonation time ratio

Phonation time ratio is the percentage of time spent speaking as a proportion of the total time taken for the speech sample (Kormos & Dénes, 2004; Towell et al., 1996). This ratio was calculated as the speaking time (excluding pauses) divided by the total

individual performance time including pauses x 100. The pre-intervention measure ($M = 28.07$, $SD = 9.77$) to post-intervention measure ($M = 30.77$, $SD = 8.51$) increase of 9.6% was not statistically significant $t(7) = -.673$, 95% CI [-12.2, 6.8], $p = .523$, $d = .29$, for Study Two participants. No comparison was possible with Study One as pauses were not able to be measured in Study One due to using MP3 audio file format, which was not compatible with PRAAT. Table 5.9 shows that for five speakers (D1, D2, C2, C1, and B1) the percentage of time spent speaking increased from pre- to post-measure.

Table 5.9 Phonation time ratio change between pre- and post-measure

ID	Pre phonation time ratio	Post phonation time ratio	Change
D1	22.09	42.9	+20.81
D2	14.11	25.99	+11.87
C2	20.77	28.24	+7.47
C1	27.73	33.42	+5.69
B1	36.49	40.88	+4.39
B2	23.79	16.68	-7.1
A2	36.61	26.16	-10.44
A1	42.95	31.9	-11.06

(all inclusive)

5.7.4 Articulation rate

Another measure of fluency is the articulation rate which shows the pace of actual speaking when pauses are excluded. It was calculated as syllables divided by speaking time (excluding pauses) x 60 (Towell et al., 1996). The pre- and post-articulation rate is shown in Table 5.10 for each speaker in Study Two. The difference between untrimmed pre-measure articulation rate ($M = 155.97$, $SD = 25.58$) and post-measure ($M = 150.99$, $SD = 22.44$) was not statistically significant, with a small effect size and confidence

intervals passing through zero $t(7) = .903$, 95% CI [-8.1, 18.0], $p = .397$, $d = .21$. From pre- to post-measure, four out of eight speakers showed a decrease in articulation rate. The decrease indicates slower articulation or formation of meaningful utterances. This slowdown suggests that more online processing is occurring during the stream of speech in the post-measure than during the pre-measure conversation. Perhaps speakers were trying to retrieve and use multi-word expressions that they had learned over the six weeks but had not yet developed automatized knowledge of. Again, comparison is not possible with Study One due to the different audio formats.

Table 5.10 Articulation rate

ID	Pre articulation rate	Post articulation rate	Change
C2	126.53	144.21	+17.69
B1	155.24	165.66	+10.42
A1	115.38	118.48	+3.1
C1	171.16	171.56	+0.4
D1	197.29	188.09	-9.2
D2	154.92	141.66	-13.26
B2	165.82	143.61	-22.21
A2	161.4	134.66	-26.74

(untrimmed syllables uttered per minute not including pauses)

5.7.5 Mean length of runs compared with Study One

A comparison of gain scores for untrimmed mean length of runs from pre- to post-measure showed no significant difference between Study One experimental group dyads ($M = -.172$, $SD = .800$) and Study Two experimental group dyads ($M = -.117$, $SD = .604$), $t(17) = -.127$, 95% CI [-.97, .86], $p = .90$, $d = .08$. The experimental group in Study One did not change significantly from pre- ($M = 3.7$, $SD = .86$) to post-measure

($M = 3.5$, $SD = .78$) $t(14) = .831$, 95% CI $[-.27, .62]$, $p = .42$, $d = .24$. Neither did Study Two group change from pre- ($M = 2.72$, $SD = .54$) to post-measure ($M = 2.60$, $SD = .15$), $t(3) = .386$, 95% CI $[-.85, 1.08]$, $p = .73$, $d = .30$. Therefore, it seems that mean length of runs is not affected by the experimental activities and Study Two results for mean length of runs are similar to those from Study One.

5.7.6 Individual change in mean length of runs for Study Two participants

The mean length of runs (total syllables divided by number of runs), for both trimmed and untrimmed transcripts can be seen in Table 5.11 for Study Two participants. The mean length of runs did not change significantly from the pre- to post-measure for untrimmed $t(7) = .707$, 95% CI $[-.36, .67]$, $p = .50$, $d = .30$, or trimmed dialogues $t(7) = .184$, 95% CI $[-.54, .63]$, $p = .86$, $d = .06$.

Table 5.11 Mean length of runs for Study Two participants

	Trimmed dialogues	Untrimmed dialogues
	<i>M (SD)</i>	<i>M (SD)</i>
Pre-measure	2.79 (.69)	2.69 (.64)
Post-measure	2.75 (.61)	2.53 (.40)

Table 5.12 shows the individual speakers' change in mean length of runs from pre- to post-measures for both the trimmed conversations and the untrimmed conversations. The final column shows untrimmed change minus trimmed change. For five speakers there is a decrease between the trimmed and untrimmed mean length of run change. This difference between mean length of run is likely due to greater use of single run fillers in the post-test.

Table 5.12 Change between pre- and post- mean length of runs

ID	Trimmed			Untrimmed			Difference between trimmed and untrimmed change
	Pre mean length of run	Post mean length of run	Change mean length of run	Pre mean length of run	Post mean length of run	Change mean length of run	
A1	3.91	2.47	-1.44	3.36	2.41	-0.95	+0.49
D2	2.29	2	-0.29	1.95	2	+0.05	+0.34
C2	1.73	2.5	+0.77	1.73	2.53	+0.81	+0.04
B2	2.5	2	-0.5	2.53	2	-0.53	-0.03
C1	2.47	2.78	+0.32	2.47	2.68	+0.21	-0.11
B1	3	3.5	+0.5	2.86	3.09	+0.23	-0.27
D1	3	3.27	+0.27	3.09	3	-0.09	-0.36
A2	3.46	3.46	0	3.54	2.57	-0.97	-0.97

For A2 (a slow speaker), who was mentioned above for using *I think I'll*, there was no change in mean length of runs in the trimmed conversation, but once fillers and repetitions were included, his mean length of run shortened for the post-test. A look at the transcripts reveals that he only used one filler in the pre-measure conversation, while he increased his use of filled pauses to four in the post-measure conversation. He also repeated his words in the post-measure conversation, whereas in the pre-measure, he made no repetitions. When filled pauses were included in the analysis, they shortened the mean length of runs because they often appeared as a single morpheme run, which reduced the average length of run.

D1 (a fast speaker), and also mentioned above, increased mean length of runs slightly in the post-measure trimmed version, but showed a very slight decrease once

fillers and repetitions were included. D1 did not use fillers in the pre-measure conversation but used five in the post-measure. For A1 (a new speaker example) the change in untrimmed mean length of runs was less of a decrease than the trimmed change. A1 used two single run fillers in the pre-measure and again in the post-measure, so there was no increase in fillers unlike A2.

In response to research question three, it appears that fluency as measured by speech rate and mean length of runs in Study Two support the same conclusions as Study One. That is, the increase in speech rate and lack of significant change in mean length of runs was similar for Study One and Study Two. The new measures of fluency added for Study Two of phonation time ratio and articulation rate showed no statistical increase as a result of the intervention.

5.8 Aural vocabulary knowledge

Research question four was about whether general aural vocabulary knowledge (not including the multi-word expressions) was impacted by the intervention. Study One showed no impact, but would Study Two support this finding? In this section, Study One and Study Two results are compared, followed by a consideration of the Study Two within group results.

5.8.1 Aural vocabulary knowledge compared with Study One

The Listening Vocabulary Levels Test (LVLТ) was included in Study One in order to see whether vocabulary knowledge untargeted by the experiment might increase, particularly in the control group (see Chapter Three for reasoning). There was no control group in Study Two. However, we can see from the following results that unrelated vocabulary cannot be expected to increase as a result of this focus on multi-word expressions.

First, it appeared that the participants in both Study One and Study Two had similar vocabulary knowledge in both the pre- and post-test. Using the Independent Samples Kruskal-Wallis test, as the Study One experimental group distribution was deviant from normal $D(46) = .140, p = .024$, no difference was found between Study

One and Study Two LVLT test results in the pre-test ($n = 118$) $\chi^2(2) = 1.79$, asymp $p = .408$, or in the post-test ($n = 99$) $\chi^2(2) = 1.48$, asymp $p = .478$.

5.8.2 Aural vocabulary knowledge for Study Two alone

Looking at the Study Two LVLT results separately, there appeared to be one outlier in the post-test results and a significant Shapiro-Wilk score ($W = .88$, $p = 0.01$). A paired samples t -test showed no significant difference between the pre- ($M = 55.6$, $SD = 5.15$) and post-test scores ($M = 56.6$, $SD = 5.30$), $t(21) = -1.2$, 95% CI $[-2.97, .785]$, $p = .240$, $d = .19$. This result was confirmed with a related samples Wilcoxon Signed Rank Test ($Z = 155.0$, asymp $p = .168$). When the outlier was removed the distribution normalized, but the pre- ($M = 55.9$, $SD = 5.1$) to post-test ($M = 57.4$, $SD = 3.9$) difference was still not significant $t(20) = -1.96$, 95% CI $[-3.2, .10]$, $p = .06$, $d = .33$. Therefore, overall, Study Two participants did not improve their general aural vocabulary knowledge.

When the levels of the LVLT were investigated separately, distribution was generally non-normal as shown by the Shapiro Wilk scores in Table 5.13, and therefore non-parametric tests were used. The increase in median score for level one was found to be significant from pre- to post-test, but the increases for level two and three were not significant.

Table 5.13 LVLT differences by level

LVLT ($n = 22$)	Shapiro Wilk stat (sig.)	Median	IQR	Wilcoxon signed rank significance
Level 1 pre	.826 (.001)*	22.0	2.5	
Level 1 post	.741 (< .001)*	23.0	2.0	.020*
Level 2 pre	.914 (.056)	19.5	3.0	
Level 2 post	.936 (.161)	20.0	4.0	.145
Level 3 pre	.884 (.015)*	14.0	3.75	
Level 3 post	.919 (.072)	15.0	4.0	.623

(*significance < .05)

Therefore, in response to research question four, experimental activities in Study Two appeared to improve aural vocabulary for the most frequent 1000 words, while no effect was seen for the most frequent 2000 and 3000 words. For comparison, the experimental group in Study One also showed no statistical change in overall aural vocabulary knowledge as a result of the treatment. These results will be discussed further in Chapter Seven.

5.9 Which activities are considered the most useful by learners?

Research question five asked which activities were considered the most useful by learners. In order to answer this question, feedback was gathered from participants regarding the perceived usefulness of each activity in Japanese through an online survey (see Chapter Three for more details). In this section, the Study Two feedback is compared with feedback from Study One. However, because I used a six-point Likert scale in Study Two rather than the five-point scale used in Study One, direct comparison is difficult. A direct comparison would have involved collapsing the Likert scales to simply agree or disagree, but collapsing the Likert scales would also involve removing the neutral data points from Study One, which I felt would distort participant feedback. Therefore, I have reported the percentage agreement in order to compare student sentiment between the two studies.

Feedback was also gathered through a focus group interview. The focus group included two female students (who will be referred to with pseudonyms: Moe, who had greater proficiency with English (she happened to teach English as a tutor for high school students) and Nene, who perhaps represented the majority of the students who had no contact with English outside of class. As in Study One, no male students volunteered for the focus group. Table 5.14 presents the activities starting with those agreed upon as useful by participants. The 6-point Likert scale described in 5.1.3 was collated to simply agree or disagree for simplicity. After the feedback on the activities, there is a final section reporting participant feedback on the three-unit themes in terms of perceived usefulness and enjoyment.

Table 5.14 Activity usefulness as agreed by participants

Activity	% responded that activity was useful
Shadowing ($n = 15$)	100
Role-play ($n = 20$)	95
Phrase instruction ($n = 20$)	90
Mingle jigsaw ($n = 20$)	90
Decreasing time role-play ($n = 20$)	90
Dictogloss ($n = 19$)	89
Free role-play ($n = 20$)	80
Record role-play ($n = 20$)	76

5.9.1 Everyone agreed that shadowing was useful

All participants in Study Two agreed that shadowing was a useful activity, which contrasts with Study One where only 69% considered shadowing useful. There was no change to the shadowing activity from Study One to Study Two. Shadowing was also considered enjoyable by 80% of participants in Study Two. One participant commented regarding shadowing in the survey, “if I continue (shadowing) it feels like I’ll get better.”

In the focus group, Nene mentioned that the audio at times was too fast for her to shadow properly and she was not able to keep up. I asked if she was able to pause the audio when she needed to, and she responded that she had not known that pausing was allowed. In contrast, Moe said that she paused the audio every time she did shadowing. I believe I did not specifically tell the students that they could pause freely, assuming that they would pause the audio when necessary since they had control of their own audio. It seems that specific instruction about pausing is necessary as students may not feel free to pause without explicit permission.

5.9.2 Role-play activity feedback results

Role-play was considered useful by 95% and enjoyable by 85%. This activity continued to be a favourite among learners as in Study One, where 73% responded that role-play was useful. One comment on the survey stated that they could feel the atmosphere of the role-play scenario, giving it a realistic vibe. Moe and Nene both agreed that it was an enjoyable activity. Moe said it was useful for pronunciation practice, and Nene said it was useful for speaking practice.

5.9.3 Phrase instruction activity feedback results

Phrase instruction was perceived by 90% to be useful, which mirrors the response from participants in Study One. Unlike Study One, though, participants were also asked whether phrase instruction was enjoyable and 75% answered that it was. Therefore, this activity appears to be appropriate for the context and learners.

5.9.4 Mingle jigsaw activity feedback results

Mingle jigsaw was considered useful by 90% of participants. This usefulness rating is an increase from only 52% in Study One. It was also considered enjoyable by 80%, which seems like an improvement from Study One, although enjoyment was not specifically measured then. In Study One, only 6% agreed that the mingle-jigsaw was too difficult, and 70% were neutral. However, in Study Two with the six-point Likert scale forcing an opinion, difficulty was felt by the majority in all aspects: socially (85%), technically (75%), and linguistically (80%).

In the focus group, Moe said that the mingle jigsaw activity was fun because it was like a game. She said that mingle jigsaw was not difficult socially because everyone wanted to collect the phrases so they were actively participating. This comment is quite different from Study One, where the focus group participants noted that there were male students who would stand and do nothing instead of actively trying to collect the phrases. Perhaps the addition of the time limit helped push the students to action.

In the Study Two focus group regarding mingle jigsaw, Nene noted that the phrases were not difficult but together with Moe, she found it difficult to remember the multi-word expressions between the verbal exchange and getting back to her desk to

write them down. Moe said that remembering the four-word expression alone was not as useful as remembering a whole sentence would be, since a sentence would have more meaning. This comment echoes Elaine from Study One, who also noted that remembering such short phrases made it hard to assign meaning to them. Many of the target multi-word expressions in this study appear incomplete in isolation, for example: *but so does she, are you going to, when you get to*. This incomplete structure likely makes them more difficult to remember when compared to structurally complete expressions such as: *have a good day* and *thank you very much*. It would be worthwhile to investigate empirically whether multi-word expressions that appear incomplete are best remembered within a sentence or as an isolated chunk. One could compare results from a mingle jigsaw done using isolated multi-word expressions that appear structurally incomplete, versus a mingle jigsaw done using the same multi-word expressions embedded in sentences. For this study however, it seems that adding the five-minute time limit to the mingle jigsaw has been effective to focus participants on the task of collecting expressions, rather than being concerned about who to speak to.

5.9.5 Decreasing time role-play activity feedback results

The decreasing time role-play was considered technically difficult by 90% of participants. In Study One, 60% responded that the decreasing time role-play was too difficult. In response to that feedback, the initial times allowed for the directions and café unit were decreased, while the time allowed for the hotel role-play was increased and the hotel role-play was shortened. Also, the target expressions were displayed on the class screens as a reference. Although Study Two participants responded that the decreasing time role-play was still difficult, 90% found it useful compared with 60% from Study One, and 95% found it enjoyable.

In the focus group, Moe said that she role-played based on the meaning rather than trying to recall word for word, but by going off script some of her partners would get stuck and not know how to respond. Nene said that she was not able to remember English and speak so freely as Moe and needed to role-play just like the model script. She said she often forgot what to say and ran out of time while trying to recall the words.

Moe also said that the time pressure made the decreasing time role-play feel like a game, where the challenge was to finish before the time ran out. However, they both agreed that the hotel role-play was the most difficult as there were many parts to remember, and they felt that no one was able to complete it within the time frames. Perhaps the hotel role-play needed a little extra scaffolding to make it more achievable and more time allowance than the slightly short role-plays.

5.9.6 Dictogloss activity feedback results

The dictogloss was considered useful by 89% of participants, an increase from only 54% in Study One. In Study One, 64% responded that the dictogloss was too difficult, in Study Two, over half of the respondents reported that dictogloss was technically and linguistically difficult. However, dictogloss was also considered to be enjoyable by 84% of Study Two participants. One survey comment complained that the speed of the audio was too fast. In the focus group, Moe and Nene also said that the speed of the audio made dictogloss difficult. Moe said, “In our group of four there was always someone who said they could not catch anything from the audio and had taken no notes, but there was also likely to be at least one or two people who had listened carefully and taken notes.” Moe also noted that dictogloss felt like the most difficult activity because it challenged them to listen and write what they heard in English. It seems that setting a time limit and working in groups improved the usefulness and enjoyment of dictogloss for the learners when contrasted with the Study One experience.

5.9.7 Free role-play activity feedback results

80% of participants in Study Two reported that the free role-play was useful (up from 61% in Study One). 75% of participants in Study Two also found the free role-play to be enjoyable. In Study One, 46% found the free role-play to be too difficult and more scaffolding was provided in response to that feedback. The free role-play was found to be linguistically difficult by 85% of the Study Two participants. It is not surprising that learners found the free role-play linguistically difficult, since this activity was the only activity in the set that required learners to role-play completely original

dialogues, challenging them to find the necessary words for the new ideas they wanted to express.

One comment on the free role-play from the survey reiterated what focus group participants had mentioned in Study One; that the success of this learning activity really depended on who a learner was partnered with. Some partners were great for adlibbing conversation, while others were unable to produce much language at all. In these classes, the partners were always randomly assigned at the beginning of each class, which meant that participants were never with one partner for very long and were able to experience a variety of speaking partners. In the focus group, Nene said, “free role-play is difficult because even though I know what I want to say, I can’t remember how to say it in English.”

5.9.8 Recording the role-play activity feedback results

Recording the role-play was considered useful by 76% of respondents, similar to Study One where 73% considered it useful. In Study One, 28% responded that recording the role-play was too difficult, 58% chose a neutral response, and only 15% responded that it was not too difficult. Study Two participants considered recording the role-play as technically difficult (81%) and linguistically difficult (76%).

In the focus group, Nene said, “recording the role-play was difficult because we could not look at notes and most people in the class wanted to role-play just as the model script, so it was difficult if you could not remember the script.” Moe said, “some people were happy to continue even if they made a mistake or went off script while others stop and say we are doing it wrong.” Nene also noted that listening back to the recording was useful for hearing the parts she was not good at. Regarding listening back, Moe said, “listening back was good, I realised that I stopped more than I thought I had.” Even though recording the role-play was perceived as technically and linguistically difficult, it seems that it was useful for self-reflection purposes.

5.9.9 Unit themes useful and enjoyable

The reader might recall that a question was added to the feedback survey about the usefulness of the unit themes. The café unit as well as the directions unit were

perceived to be useful for the future by 90% of participants ($n = 20$), while the hotel theme was perceived to be useful by 85%. The directions and hotel units were considered enjoyable by 80% of respondents and the café unit was considered enjoyable by 75% of respondents. Based on this feedback, the unit themes appear to be appropriate for these participants.

5.10 Study Two results summary

In this chapter, I have justified and described changes to the methodology for Study Two in line with action research methods. I have compared the results of Study Two with Study One. The results suggest that Study Two was no more or less effective than Study One for 1) building knowledge of the target multi-word expressions, 2) enabling voluntary use in conversation and 3) improving speaking fluency. The results also show 4) no effect on wider vocabulary learning. The final research question was perhaps the most important for this study; to discover the perceived usefulness of activities by learners. It was important to discover whether the changes made to the activities were worthwhile, and it seems that the changes did improve the usefulness and enjoyability of the classroom activities from the learners' perspective. Teacher practitioners are often keen to hear about recommended classroom activities. This study provides evidence for teacher practitioners that shadowing is indeed considered useful by (these) learners, with 100% of participants agreeing that shadowing was a useful activity for language learning.

Reflecting on results from both Study One and Study Two, it seems clear that repeated use of target multi-word expressions through varied interactive activities increases productive form meaning links with multi-word expressions. There is a suggestion of a link between the activities and increased fluency, but more evidence is definitely required to determine any relationship. Also, the activities have yet to increase productive and voluntary use of target multi-word expressions in conversation. In order to investigate the relationship between the activities and multi-word expression productive ability and fluency, I felt it necessary to conduct a third study that essentially replicated the methods of Study Two. In Chapter Six, a replication study is described

which includes more participants and a control group. Following on from that, Chapter Seven draws all results and observations together in an overall discussion.

6 Chapter Six: Study Three method and results

Following the action research paradigm of this project, this chapter builds on the insights and results from Study One and Study Two and serves as a replication of the methods from Study Two. Study Two lacked a control group due to scheduling limitations. Study Three includes a control group and expands the participant pool in the hope of producing more generalizable results than were able to be produced from Study Two. I made slight adjustments to the classroom activities for Study Three, so these will first be described and justified. The research questions will then be introduced, followed by participant details, results and participant feedback.

6.1 *Changes made to improve classroom activities*

A few minor adjustments were made to the classroom activities for Study Three; I will now describe these in the order shown in Table 6.1. As the classroom teacher, I considered the feedback received from Study Two participants through the online survey and focus group. In consideration of the feedback, I decided to make three minor changes to improve the classroom activities in Study Three.

Table 6.1 Changes made to experimental classroom activities for Study Three

1	Shadowing: Add explicit instruction allowing pausing
2	Hotel decreasing time role-play: Provide booking form realia
3	Hotel decreasing time role-play: Increase time allowed by 20 seconds

6.1.1 *Shadowing: Add explicit instruction to allow pausing*

The Study Two focus group discussion revealed that some participants felt free to pause and replay the audio during shadowing, while others had not realized that they could. I had assumed that participants would pause and play as they desired since they had the controls. However, the focus group feedback revealed that some participants needed to be told that they were free to pause and play as they liked. Therefore, in Study Three, I gave explicit instruction to participants that they could pause and replay difficult parts of the audio during the shadowing activity as they wanted to.

6.1.2 Hotel unit decreasing time role-play: Provide booking form realia

Despite adjustments made to the model hotel role-play in Study Two, feedback from participants still singled it out as being more difficult than the café and directions role-plays. Therefore, I created hotel booking form realia for staff role-players to fill in with guest information during the role-play just as a hotel worker might do. The booking form reminded them about which information to elicit. This addition of a booking form was intended to simulate the hotel booking situation more realistically and provided extra scaffolding. See Appendix 26 for the hotel booking form realia.

6.1.3 Hotel unit decreasing time role-play: Increase time

Despite shortening the model hotel role-play in Study Two, focus group participants remarked that almost no one was able to complete the hotel dialogue within the time allowed for the decreasing time role-play activity. Therefore, I decided to increase the time allowed by 20 seconds for each of the three time periods for the hotel unit. The increase was only for the hotel unit since it was longer than the café and directions units. So, the first time period started with 2 minutes and dropped down to 1.4 minutes and finally 1.2 minutes.

The above adjustments to the classroom activities were made in order to improve the classroom experience and effectiveness of the activities in Study Three. While the adjustments were minor, they are recorded here for the reference of other practitioners in order that we may all learn from this study and to enable close replication of this method in the future. Now that I have described the adjustments to the method for Study Three, I will introduce the research questions, participants and results.

6.2 Study Three

Study Three is essentially a replication of Study Two. The same materials were used for all tests and class instruction. A control group using the same course as was used in Study One was included so that comparison between the intervention and linked skills classes could be made.

6.2.1 Research questions

- 1) Is the increase in knowledge of target meaning and form found in previous studies replicable?
- 2) Does practice of target multi-word expressions through different modes increase voluntary use in conversation?
- 3) What does the intervention reveal about the relationship between use of multi-word expressions and fluency?
- 4) Does greater vocabulary knowledge correlate positively with multi-word expression learning gain? Does focus on multi-word expressions affect wider vocabulary knowledge?
- 5) Which activities are considered most useful by the learners?

6.2.2 Study Three participants

As in the previous two studies, participants were recruited from the same engineering university in northern Japan. Only data from those who consented to participate are included in the analysis and reporting. Participants self-reported their most recent TOEIC score range. The majority claimed TOEIC scores ranging from 255 to 400 (roughly equivalent to A2 CEFR), as seen in Table 6.2. For comparison with Study One and Study Two, see Chapter Four (4.1.1) and Chapter Five (5.4.2) respectively.

Table 6.2 Study Three participants' TOEIC score range percentages

TOEIC score	Experimental (<i>n</i> = 65)	Control (<i>n</i> = 51)
10-250	7.7%	7.8%
255-400	53.8%	58.8%
405-600	36.9%	23.5%
605-780	1.5%	9.8%

Non-normal data distributions were identified with significant Shapiro-Wilk scores in both the experimental group ($W = .795, p < 0.01$) and the control group ($W = .805, p < 0.01$). Therefore, a Mann-Whitney U test was used to compare the groups' proficiencies based on TOEIC scores, finding no significant difference between them: $U = 1,683.5, z = .162, \text{asympt } p = .871, r = .015$. Based on these score ranges, English proficiency appeared to be roughly comparable across the experimental and control groups. The TOEIC score ranges correlated positively with the pre-intervention Listening Vocabulary Levels Test results $rs(116) = .56, p < .001$. The TOEIC score range also correlated positively with the pre-intervention MWE cloze test $rs(116) = .40, p < .001$. Meanwhile, the LVLT test score correlated positively with the MWE cloze test score $rs(116) = .47, p < .001$. The link between vocabulary knowledge and proficiency is well-documented and so we would expect to see these correlations (Stæhr, 2008).

Of 116 participants, 101 (87.1%) were male. The breakdown of gender and course of study (department) is shown in Table 6.3. The experimental group was made up of a single major (Applied Sciences), whereas the control group was a mix of two majors (Civil Engineering and Mechanical, and Aerospace and Materials Engineering). Both groups had to take two compulsory English classes to fulfil their graduation requirements. There were also three international students in the control group from China, while there were no international students in the experimental group.

Table 6.3 Gender and major for Study Three participants

	Experimental ($n = 65$)	Control ($n = 51$)
Male	54 (83.1%)	47 (92.2%)
Female	15 (12.9%)	4 (7.8%)
Applied Sciences	65 (100%)	0
Civil Engineering	0	20 (39.2%)
Mechanical, Aerospace and Materials Engineering	0	31 (60.8%)

6.2.3 Measuring effect

Following the same method as Study Two, participants sat pre- and post-intervention tests and recorded a dialogue which measured their knowledge and use of target multi-word expressions, along with speaking fluency. Table 6.4 shows the seven measures used to compare participant knowledge along with the four main purposes of the measures. An alpha level of 0.5 was used to identify statistical significance. Distributions were checked for normality and are mentioned only when there was a significant deviation from a normal distribution. I will report the results in the same order as presented in Table 6.4.

Table 6.4 Measures used in Study Three with their purposes

Measure	Purpose
Multi-word expression cloze test scores	Measure productive form and meaning knowledge of 30 multi-word expressions
Use of multi-word expressions in conversation	Measure ability to use multi-word expressions (procedural knowledge)
Speech rate	Measure spoken fluency (automaticity)
Phonation time ratio	
Articulation rate	
Mean length of runs	
Listening Vocabulary Levels Test	Measure aural vocabulary knowledge (not including the 30 multi-word expressions)

6.3 Multi-word expression cloze test results

In order to investigate the first research question, the same pre- and post-intervention multi-word expression cloze test was used as in Study One and Study Two. For details on this test please refer to the methodology in Chapter Three. Only those who were present for both the pre- and post-test are included in this analysis. The pre-test distribution was deviant from normal for the control group with a significant Shapiro-Wilk score $W = .899$, $p < .001$. A Mann Whitney U test showed that the control and experimental groups were starting with comparable knowledge of the target multi-word

expressions in the pre-test $U = 2,005.5$, $z = 1.94$, asymp $p = .053$, $r = .18$.

For the post-test, the control group distribution was again deviant from normal $W = .935$, $p = .008$. A Mann Whitney U test showed that the control and experimental groups had significantly different scores: $U = 2,762.0$, $z = 6.14$, asymp $p < .001$, $r = .57$. A comparison of gain scores showed a significant difference between the control group and experimental group with a large effect size, $t(114) = -9.93$, $p < .001$, $d = 1.88$. We can be 95% sure, based on the confidence intervals, that the mean difference between the two groups lies somewhere between -7.44 and -4.97 . Since the confidence intervals do not pass through zero, the difference between the groups is statistically significant. Table 6.5 shows the descriptive statistics for the multi-word expression cloze tests (pre- and post-intervention), and the average gain for each group.

Table 6.5 Descriptive statistics for Study Three multi-word expression cloze test

	Pre <i>Mdn</i>	Post <i>Mdn</i>	Gain <i>M (SD)</i>
Experimental group ($n = 65$)	<i>Mdn</i> = 11.45	<i>Mdn</i> = 20.17	$M = 8.71$, $SD = 3.64$
Control group ($n = 51$)	<i>Mdn</i> = 10.47	<i>Mdn</i> = 12.97	$M = 2.50$, $SD = 2.92$

6.3.1 Replicable effect

The Study Three results can be compared with Study Two and Study One results as the test was the same. In Study One, the difference with the control group had an effect size of $d = 1.41$. As detailed above, the Study Three effect size was even larger $d = 1.88$. When the change between the pre- and post-test knowledge of the target multi-word expressions for the experimental group alone in Study Three is compared, there is a significant difference with a large effect: $t(64) = -19.31$, 95% CI $[-9.61, -7.81]$, $p < .001$, $d = 1.89$. In Study Two, the effect size was also large: $d = 1.89$. In Study One also, the within group comparison effect size was large: $d = 1.74$. These three experiments have shown and confirmed (through replication) that a classroom focus on multi-word expressions results in increased productive knowledge of form and meaning.

Whether such increased knowledge is accompanied by increased use of the multi-word expressions is the topic of the next section.

6.4 Use of multi-word expressions in conversation

In order to answer the second research question regarding whether practice leads to an increase in use of target multi-word expressions, pre- and post-intervention dialogues were recorded and analysed using the same methods as described in Chapter Five. In Study Three, there were 24 speakers from the experimental group and 22 speakers from the control group (limited to those who gave consent, spoke with the same partner in the same role for both the pre- and post-intervention dialogue). Many participants were not included in this analysis because they spoke in the opposite role from pre- to post-test, and others had different partners due to some participants being absent on the test day. A list of multi-word expressions that were counted from the Study Three pre- and post-intervention dialogues can be seen in Appendix 27. In the next section I will report analysis of the use of multi-word expressions in the pre- and post-intervention dialogues.

6.4.1 Between group comparisons of multi-word expression use

Gain scores for multi-word expression use were calculated, and distribution was found to be deviant from normal in both the experimental group ($D(24) = .24, p = .001$) and the control group ($D(22) = .24, p = .002$). Therefore, a Mann-Whitney U test was used to compare the change in use of multi-word expressions across the two groups. A significant difference with a large effect was present for the change in multi-word expression use from pre- to post-measure between the experimental and the control group: $U = 429.5, z = 3.66, \text{Asymp. } p < .001, r = 0.54$. Further analysis using a non-parametric paired samples t -test (related-samples Wilcoxon signed rank test) to compare changes within the groups showed a significant change within the experimental group from pre- to post-measure with a large effect size for words used in target multi-word expressions: $Z = 289.5, \text{asympt } p < .001, r = 0.58$. In contrast, the control group did not show a difference from pre- to post-measure: $Z = 69.0, \text{asympt } p = .469, r = 0.15$. Table 6.6 shows the descriptive statistics for multi-word expression use in the recorded

dialogues (pre- and post-measure), and the average gain in use for each group. In the next section, I will compare Study Three multi-word expression use with Study One.

Table 6.6 Descriptive statistics for Study Three multi-word expression use

	Pre <i>Mdn</i>	Post <i>Mdn</i>	Gain <i>M (SD)</i>
Experimental group (<i>n</i> = 24)	<i>Mdn</i> = 6.0	<i>Mdn</i> = 9.5	<i>Mdn</i> = 3.5
Control group (<i>n</i> = 22)	<i>Mdn</i> = 4.5	<i>Mdn</i> = 4.5	<i>Mdn</i> = 0

6.4.2 Comparing multi-word expression use with previous studies (by dyad)

In Study Three, there was a clear increase in the voluntary use of multi-word expressions in conversation by the experimental group. We can use effect sizes to compare the effect of the experiment on target multi-word expression use between the three studies. Keep in mind that the role-play prompt was different for Study One, and that analysis was done by dyads rather than individual speakers. In Study One, while there was no meaningful effect based on group in the pre-test ($d = .224$), the post-test showed a medium effect based on group ($d = .70$). For Study Three, dyad averages for multi-word expression use are shown in Table 6.7. No significant difference was found between the experimental and control group in the pre-test; $t(20) = .268$, 95% CI [-4.74, 6.14], $p = .791$, $d = .011$ for use of target multi-word expressions. However, there was a significant difference with a large effect size between the groups in the post-test; $t(17.70) = -2.16$, $p = .045$, 95% CI [-18.05, -.248], $d = .90$, but note the df is lower as equal variances are not assumed. The greater effect in Study Three (compared with Study One) could perhaps be associated with the improvements made to the classroom activities.

Table 6.7 Descriptive statistics for Study Three multi-word expression use (by dyad)

	Pre <i>M</i> (<i>SD</i>)	Post <i>M</i> (<i>SD</i>)	Gain <i>M</i> (<i>SD</i>)
Experimental group (<i>n</i> = 12)	<i>M</i> = 11.50 <i>SD</i> = 4.87	<i>M</i> = 22.75 <i>SD</i> = 12.52	<i>M</i> = 11.25 <i>SD</i> = 10.75
Control group (<i>n</i> = 10)	<i>M</i> = 12.2 <i>SD</i> = 7.32	<i>M</i> = 13.6 <i>SD</i> = 6.96	<i>M</i> = 1.4 <i>SD</i> = 6.60

Comparing the increase in multi-word expression use within the experimental group in each of the three studies, we can see that the effect size is large. For Study One: $d = .81$, for Study Two: $d = .94$, and for Study Three: $t(11) = -3.6$, $p = .004$, 95% CI [-18.08, -4.42], $d = 1.18$. Combined, these studies provide evidence that the practice of multi-word expressions through different modes using classroom activities is effective for increasing voluntary use of target expressions in conversation. As mentioned earlier, the prompt for the pre- and post-test role-play dialogue was changed between Study One and Study Two with the intention to expand the possible target multi-word expressions that might be used. The next section compares the range of multi-word expressions used between the three studies, to see if the change in prompt did in fact expand the range of multi-word expressions used.

6.4.3 Comparing range of multi-word expression use

In Study Three, a range of seven target multi-word expressions were used in their complete target form in 15 instances, while a further range of 15 target multi-word expressions were partially used in the post-measure dialogues. Therefore of the 30 target multi-word expressions, 22 were fully or partially used in the post-intervention dialogues by the experimental participants. In Study Two, only nine expressions were fully or partially used. Keep in mind that only four dialogue pairs were analysed in Study Two, which limits the generalizability of those results. In Study One, when the role-play instructions were limited to a single scenario, only three target expressions were used in their complete form, while another nine target multi-word expressions were partially used. Therefore, based on the Study Three results, it appears that the expansion of the role-play

scenario allowed a greater range of expressions to be used.

Some expressions overlapped, and it is notable that the expressions that were used in their full form (all four words) often included words that appeared in other multi-word expressions. Bigrams and trigrams that reoccurred in the target expressions were encountered more, so it is logical that uptake of these combinations would be greater. Table 6.8 shows the instances in each study of complete use of target multi-word expressions, note ‘what would you like (to)’ was used as a combined five-word expression and is therefore counted as such to avoid double counting.

Table 6.8 Post-intervention use of complete expressions

Target expression	Unit	Study One	Study Two	Study Three
I think I will	Café		1	
I will have the	Café			1
What would you like (to) combination*	Café/(Hotel)	5*		2*
What would you like	Café			2
Would you like to	Hotel	2		5
What are you going	Café			
Are you going to	Café			
How do I get	Directions			1
Until I get to	Directions			
Have a good day	Directions	1		
Thank you very much	Hotel			1
Can you tell me	Directions		1	1
When you get to	Directions			2

(words that appear in other target expressions are shown in **bold font**)

* Combination of two MWEs (not counted separately for range)

Research has linked the use of multi-word expressions with greater fluency, so the next section investigates whether greater use of the expressions lead to improvements in fluency.

6.5 Spoken fluency

Research question three asked what further light this intervention would shed on the relationship between use of multi-word expressions and fluency. In order to answer this question, four measures of fluency were investigated; speech rate, phonation time ratio, articulation rate and mean length of runs. Speech rate was investigated first.

6.5.1 Speech rate

Table 6.9 shows the mean and standard deviation for the experimental and control groups for untrimmed speech rate, when transition time is included as pausing time. There was no significant difference based on group in the pre-measure: $t(44) = .74$, $p = .46$, 95% CI [-7.49, 16.19], $d = .22$. Likewise, there was no difference based on group for the post-measure: $t(44) = .072$, $p = .94$, 95% CI [-12.50, 13.43], $d = .02$.

Table 6.9 Descriptive statistics for untrimmed speech rate

	Pre M (SD)	Post M (SD)	Gain
Experimental group $n = 24$	$M = 60.81$, $SD = 21.05$	$M = 68.46$, $SD = 22.17$	$M = 7.65$, $SD = 16.41$, $Mdn = 7.85$
Control group $n = 22$	$M = 65.16$, $SD = 18.56$	$M = 68.92$, $SD = 21.39$	$M = 3.76$, $SD = 16.25$, $Mdn = 6.28$

(transition time as pausing time)

Gain scores also showed no difference between the experimental and control groups, $t(44) = -.81$, $p = .42$, 95% CI [-13.60, 5.83], $d = .24$. This result was confirmed with a Mann-Whitney test, $U = 300.0$, $z = .792$, *Asymp. p* = .429, $r = .12$ since the control distribution was non-normal ($D(22) = .19$, $p = .036$). Therefore, it seems that there was no statistically significant difference in speech rate between the experimental and control group as a result of the intervention. However, when the experimental group speech rate was analysed internally, a statistically significant difference with a small to medium effect was evident from pre- to post-measure: $t(23) = -2.28$, $p = .03$, 95% CI [-14.58, -.72], $d = .35$. No significant difference was evident within the control group:

$t(21) = -1.1, p = .29, 95\% \text{ CI } [-10.97, 3.44], d = .19$. The next section compares the gains in speech rate for the experimental groups across the three studies.

6.5.1.1 Comparison of gains in speech rate across the three studies (by dyad)

The gains in speech rate can be compared across the three studies for the experimental group to see if the adjustments made to the method and activities had any effect on the outcome, and also to see if the results from the third study support the previous results. In Study One, within the experimental group the untrimmed speech rate increased with a medium effect size of $d = .68$. In Study Two the effect size was similar: $d = .64$. In Study Three the effect size was slightly smaller $t(11) = -1.68, p = .122, 95\% \text{ CI } [-17.08, 2.31], d = .40$. A direct comparison using a one-way ANOVA showed no significant difference between the three experimental groups for gain scores: $F(2,28) = .303, p = .741$.

When the comparisons with a control group are taken into consideration, Study One had a medium effect for the speech rate gain score difference between the experimental and control group $t(21) = 1.3, p = .229, d = .51$. However, in Study Three the difference between the gain scores for the experimental group ($M = 7.38, SD = 15.27, Mdn = 4.94$), and control group ($M = 3.95, SD = 16.15, Mdn = 7.61$), $t(20) = -.512, p = .61, 95\% \text{ CI } [-17.43, 10.56], d = .22$ by dyad was smaller. The lack of difference was confirmed with a Mann Whitney test as the control group had a deviant distribution $U = 59.0, z = -.066, Asymp. p = .974, r = .01$. Therefore, the adjustments made to methods and activities do not appear to affect the speech rate gain, even though there was a difference in multi-word expression use. The next section investigates the relationship between speech rate and multi-word expression use.

6.5.1.2 Correlation between speech rate and multi-word expression use

Regardless of group, was there an association between speech rate and multi-word expression use in the final measurement? Spearman's Rho non-parametric correlation was used as there was a non-normal distribution in the multi-word expression use measure: $D(46) = .19, p < .001$. As Figure 6.1 shows, a moderate correlation can be seen between multi-word expression use and speech rate in the post-measure, $r_s(46) = .39, p = .008$. Therefore, the first measure of fluency, speech rate, increased but not

more than the control in response to the intervention, and a moderate correlation was observed with the use of multi-word expressions. The next fluency measure that we will investigate is phonation time-ratio.

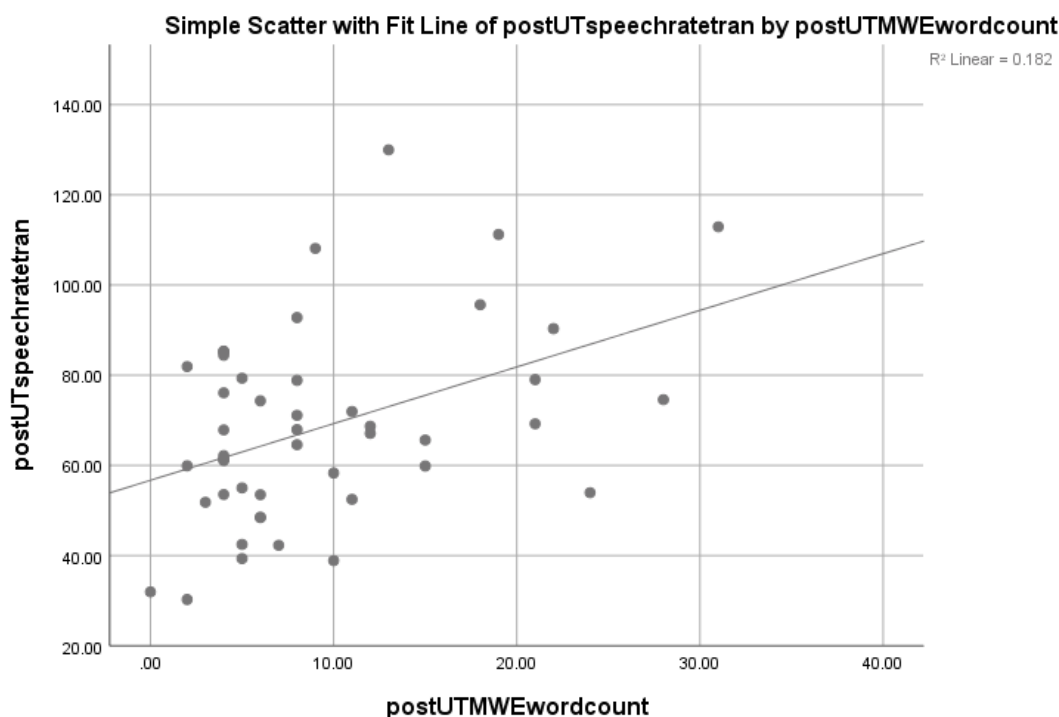


Figure 6.1 Correlation between speech rate and multi-word expression use

6.5.2 Phonation time ratio

The percentage of time spent speaking as a proportion of total time taken for a speech sample is the phonation time ratio. All pause time (including transition time) is included in the calculation. Table 6.10 shows the descriptive statistics for phonation time ratio; the experimental group had a lower phonation time ratio average in the pre-measure. The difference between the two groups in the pre-measure was not significant and the confidence intervals pass through zero, however there was a small to medium effect size: $t(44) = 1.43$, $p = .16$, 95% CI [-1.44, 8.55], $d = .42$. Again, in the post-measure, there was no statistical difference between the experimental and control groups. $t(44) = .09$, $p = .92$, 95% CI [-6.07, 6.65], $d = .03$. This lack of difference was confirmed with a Mann-Whitney U test as the data from the experimental group had a non-normal distribution $D(24) = .21$, $p = .008$; $U = 231.0$, $z = -.726$, *Asymp.* $p = .468$, r

= .10. The gain in phonation time ratio correlated positively with the gain in speech rate $r(46) = .731, p < .001$.

Table 6.10 Descriptive statistics for phonation time ratio in Study Three

	Pre <i>M</i> (<i>SD</i>)	Post <i>M</i> (<i>SD</i>)	Gain
Experimental group, <i>n</i> = 24	<i>M</i> = 33.61 <i>SD</i> = 7.76	<i>M</i> = 39.39 <i>SD</i> = 10.95	<i>M</i> = 5.77 <i>SD</i> = 9.23
Control group, <i>n</i> = 22	<i>M</i> = 37.17 <i>SD</i> = 9.05	<i>M</i> = 39.68 <i>SD</i> = 10.40	<i>M</i> = 2.51 <i>SD</i> = 5.86

Gain scores for phonation time ratio from pre- to post-measure were also calculated. No significant difference with small to medium effect emerged between the experimental and control groups, equal variances not assumed: $t(39.33) = -1.44, p = .157, 95\% \text{ CI } [-7.8, 1.3], d = .42$.

The boxplot in Figure 6.2 shows the experimental group with a wider and more positive distribution of gain for the percentage of time spent speaking. It should be noted that internally (within group), the increase in proportion of time spent speaking by the experimental group $t(23) = -3.1, p = .005, 95\% \text{ CI } [-9.7, -1.9], d = .61$ was significant with a medium effect size, while for the control group $t(21) = -2.0, p = .057, 95\% \text{ CI } [-5.1, .09], d = .26$ it was not. This result means there was more speaking and less pausing occurring in the experimental group in the post-test when compared with the pre-test. For comparison, the Study Two experimental group did not show a significant change in phonation time ratio ($d = .29$). When all the Study Three participant data was pooled together from the post-measure, only a mild correlation was seen between multi-word expression use and phonation time ratio; $rs(46) = .32, p = .03$.

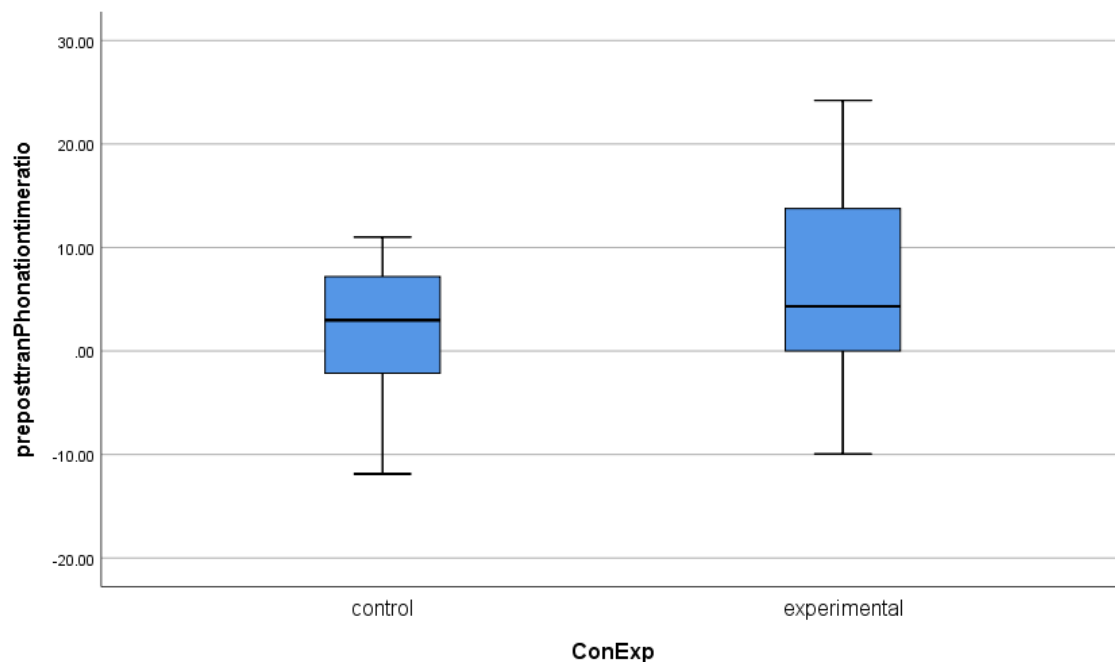


Figure 6.2 Box plot for phonation time ratio gain across groups

6.5.3 Articulation rate

Articulation rate is similar to speech rate except that pause time is excluded. The total syllables spoken are divided by the speaking time x 60, so articulation rate represents syllables spoken per minute. Table 6.11 shows the descriptive statistics for articulation rate by group. Looking at the pre-intervention articulation rate, there was no significant difference between the experimental and control groups: $t(44) = -.35$, $p = .73$, 95% CI [-19.41, 13.62], $d = .10$. There was also no significant difference in the post-intervention articulation rate between the experimental and control groups: $t(44) = .034$, $p = .97$, 95% CI [-16.8, 17.4], $d = .001$. Gain scores were compared, and no significant difference was evident $t(44) = .41$, $p = .69$, 95% CI [-12.6, 19.0], $d = .12$. The gain in articulation rate from pre- to post-intervention measures correlated positively with the gain in speech rate $r(46) = .412$, $p = .004$.

Table 6.11 Descriptive statistics for articulation rate for Study Three by group

	Pre <i>M</i> (<i>SD</i>)	Post <i>M</i> (<i>SD</i>)	Gain
Experimental group, <i>n</i> = 24	<i>M</i> = 178.15 <i>SD</i> = 29.61	<i>M</i> = 174.0 <i>SD</i> = 31.37	<i>M</i> = -4.15 <i>SD</i> = 26.42
Control group, <i>n</i> = 22	<i>M</i> = 175.26 <i>SD</i> = 25.60	<i>M</i> = 174.29 <i>SD</i> = 25.70	<i>M</i> = -.97 <i>SD</i> = 26.62

Within group analysis revealed no significant pre- to post-intervention articulation rate differences within either the experimental group, $t(23) = .770$, $p = .45$, 95% CI [-7.0, 15.3], $d = .14$ or the control group: $t(21) = .172$, $p = .87$, 95% CI [-10.8, 12.8], $d = .04$. It is worth noting that the average post-articulation rate was in fact lower than the starting articulation rate for both groups. For comparison in Study Two, the within group experimental group change for articulation rate was also negligible $d = .21$. Correlation between articulation rate and multi-word expression use was very mild in the post-measure when all participants were considered together: $rs(46) = .24$, $p = .11$.

Articulation rate unlike speech rate and phonation time ratio (which do include pausing time) did not increase. On the contrary, the articulation rate actually decreased (though not significantly). The increase in speech rate and phonation time ratio combined with no change in the articulation rate suggests that speakers were pausing less but simultaneously spreading their words out more (perhaps an indication of slower online processing). Lexical retrieval is known to put considerable demands on working memory, therefore as the learners' familiarity with the words grows, an increase in speed of lexical retrieval (in this case; expressions) could be expected (Snellings et al., 2002). The final measure of fluency for this study was mean length of runs, which we investigate in the next section.

6.5.4 Mean length of runs (untrimmed)

The mean length of runs is the average number of syllables produced in between pauses of 0.25 seconds and above. Greater mean length of run has been associated with greater automatization (Skehan, 2009; Towell et al., 1996). Table 6.12 displays descriptive statistics for the mean length of runs for the two groups. The control group

had a non-normal distribution for the pre-measure $D(22) = .24, p = .002$, so the *Mdn* is also displayed and a Mann-Whitney U test was used to compare the two groups who started with a comparable mean length of runs $U = 289.0, z = .55, \text{Asymp. } p = .58, r = .08$. The two groups also finished with a comparable mean length of runs: $t(44) = -.288, p = .78, 95\% \text{ CI } [-.41, .31], d = 0.8$.

Table 6.12 Descriptive statistics for mean length of runs in Study Three

	Pre <i>M</i> (<i>SD</i>)	Post <i>M</i> (<i>SD</i>)	Gain
Experimental group, $n = 24$	$M = 2.94$ $SD = .65$ $Mdn = 2.78$	$M = 2.84$ $SD = .58$	$Mdn = -.02$
Control group, $n = 22$	$M = 2.88$ $SD = .75$ $Mdn = 2.71$	$M = 2.79$ $SD = .63$	$Mdn = -.09$

The control group had a non-normal distribution also for the gain score $D(22) = .20, p = .018$. The gain from pre- to post-measure was negligible for both groups, with no statistical difference between the groups $U = 262.0, z = -.044, \text{Asymp. } p = .96, r = .001$. Within groups, there was no significant difference from pre- to post-intervention mean length of runs for either the experimental group $t(23) = .824, p = .42, d = .16, 95\% \text{ CI } [-.15, .34]$ or the control group $t(21) = .766, p = .45, 95\% \text{ CI } [-.16, .34], d = .13$. Since the control group had a non-normal distribution in the pre-measure, this result was confirmed with a Mann-Whitney test: $U = 113.0, z = -.44, \text{Asymp. } p = .66, r = .09$.

6.5.4.1 Mean length of runs (by dyad) compared with previous studies

In order to compare the effect of the experiment on mean length of runs with Study One, analysis was also done by dyad. Table 6.13 shows the descriptive statistics by dyad for mean length of runs for Study Three participants. Study Three showed no significant difference between the groups' gains when analysed by dyad, equal variances not assumed: $t(16.44) = 1.02, p = .32, 95\% \text{ CI } [-.18, .54], d = .41$.

Table 6.13 Descriptive statistics for mean length of runs by dyad in Study Three

	Pre <i>M</i> (<i>SD</i>)	Post <i>M</i> (<i>SD</i>)	Gain
Experimental group, <i>n</i> = 12	<i>M</i> = 2.97	<i>M</i> = 2.83	<i>M</i> = -.13
	<i>SD</i> = .53	<i>SD</i> = .41	<i>SD</i> = .53
	<i>Mdn</i> = 2.84	<i>Mdn</i> = 2.81	<i>Mdn</i> = -.10
Control group, <i>n</i> = 10	<i>M</i> = 2.73	<i>M</i> = 2.77	<i>M</i> = .04
	<i>SD</i> = .44	<i>SD</i> = .47	<i>SD</i> = .26
	<i>Mdn</i> = 2.59	<i>Mdn</i> = 2.69	<i>Mdn</i> = .09

Study One also showed a decrease for untrimmed mean length of runs by dyad, with a medium effect between the experimental and control group ($r = 0.39$). In Study Two there was no control group, so it cannot be used for comparison.

6.5.4.2 Correlation between mean length of runs and MWE use

Of the fluency measures, untrimmed mean length of runs appeared to show the strongest correlation with untrimmed multi-word expression use (word count) in the post-measure, using Spearman's rho correlation coefficient: $rs(46) = .51, p < .001$. Figure 6.3 shows the mean length of runs on the vertical axis and multi-word expression use on the horizontal axis. Those speakers who were using more words within target multi-word expressions tended to also have a longer mean length of runs. There was also a positive Pearson correlation between speech rate gain and mean length of runs gain: $r(24) = .73, p < .001$.

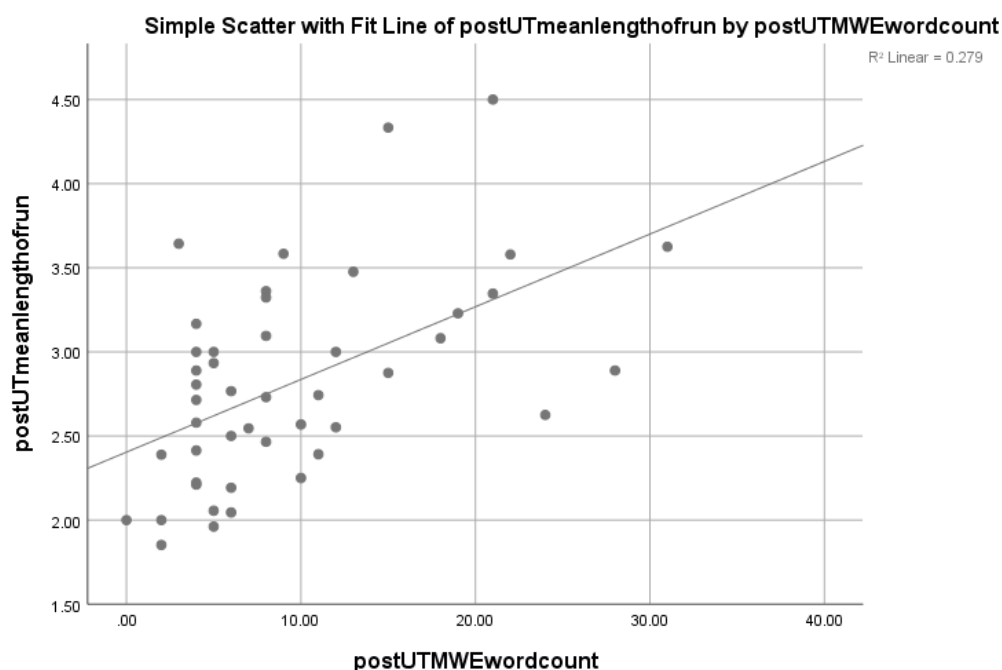


Figure 6.3 Correlation between mean length of runs and multi-word expression use

6.5.5 Summary regarding fluency outcomes

This section investigated whether fluency activities focussed on multi-word expressions improved speaking fluency. Measures of fluency investigated were speech rate, phonation time ratio, articulation rate and mean length of runs. While there appears to be no difference between the experimental and control groups in terms of speech rate, there was an increase within the experimental group suggesting there are some fluency benefits from the experimental classroom activities. While the experimental group significantly increased their phonation time ratio, there was no difference between the groups. Articulation rate did not appear to change in any discernible pattern. Mean length of runs also does not appear to be affected. The strongest correlation between multi-word expression use and fluency was shown with mean length of runs followed by speech rate.

In 6.3 an increase in form and meaning knowledge for target multi-word expressions was observed. There was also an increase in use of multi-word expressions in conversation (refer to 6.4), but at this stage of learning no strong link was seen with speaking fluency when compared with the control group, though there were some gains

at the within group level. In the next section, I report the results from the Listening Levels Vocabulary Test. Perhaps the control group might improve more than the experimental group on an external vocabulary test.

6.6 Aural vocabulary knowledge

Proficiency has correlations with vocabulary knowledge. Based on the idea of chunking (N. C. Ellis, 2001) whereby learners build chunks from the knowledge that they already have, we could expect those learners with greater vocabulary knowledge to be able to make associations between the words in the expressions (that they likely already know as individual chunks) at a faster rate than those with less vocabulary knowledge. For example, a learner who already knows the words *I*, *will*, and *think* should be faster at associating these words into the higher order chunk of '*I think I will*' than for example a learner who knows the word *I* but not *think* or *will*. Therefore, it is conceivable that we might see positive correlations between proficiency (measured by the pre-intervention Listening Vocabulary Levels Test) and improvement in multi-word expression knowledge after the intervention. The LVLT score had a non-normal distribution $D(116) = .155, p = .001$, and therefore Spearman's Rho correlation was used to investigate the relationship between the pre-intervention LVLT score and gain on the multi-word expression cloze test. A positive correlation was found: $rs(116) = .392, p < .001$. There was also a positive correlation (though not as strong) between the original LVLT score and multi-word expression use (based on count of words used within a multi-word expression) gain score for the speaking analysis $rs(46) = .293, p = .048$. When proficiency was controlled for, the differences between the groups were not changed in any meaningful way; multi-word expression use gain ($p = .004$), multi-word expression cloze gain ($p < .001$), gain in speech rate ($p = .762$).

Gain scores were calculated from pre- to post-test for the LVLT test. Outliers were investigated; one outlier from the experimental group was removed, as they had not answered some level two questions, and two outliers from the control group were removed as they had not answered some level one questions. The four remaining outliers appeared to be legitimate; three outliers in the control group and one outlier in the experimental group as can be seen in Figure 6.4. The distribution was still not

normal, with significant Kolmogorov-Smirnov scores for both the experimental group ($n = 65$, $D(65) = .130$, $p = .008$) and the control group ($n = 51$, $D(51) = .185$, $p < .001$).

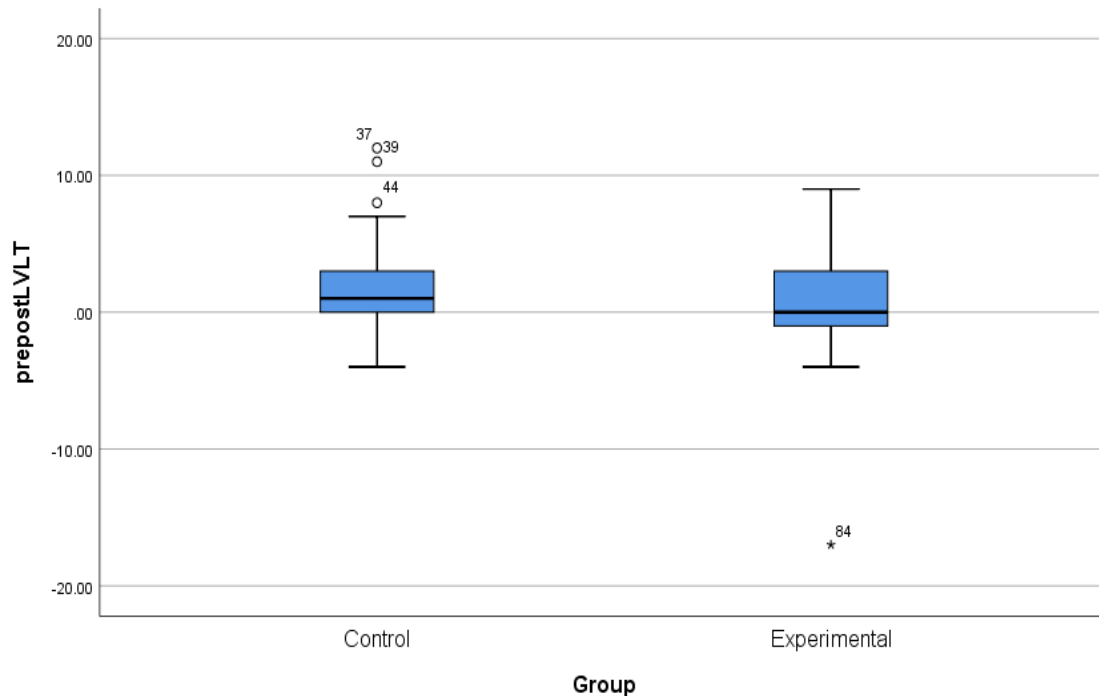


Figure 6.4 LVL T gain score comparison between groups

An independent samples Mann-Whitney U test was approaching significant difference, though the effect size remained small between the experimental group's overall gain scores ($Mdn = 0.0$) and the control group's ($Mdn = 1.0$), $U = 1,308$, asymp $p = .050$, $z = -1.957$, $r = .182$ (asymp p is used as there were ties (same values) in the data). The narrow focus of the experimental group class materials did not appear to be detrimental to their overall vocabulary learning when compared with the control group. This result supports the findings from Study One which also showed no difference between the gain scores of the experimental and control group (Study One $r = .012$). It also supports results from Study Two, where there was no control group, but no significant difference was found between pre- and post-test scores (Study Two $d = .33$).

In Study One, the control group significantly improved on level two of the test (within group comparison) and in Study Two the experimental group improved on level one. When the groups were compared separately by level for Study Three, there was a non-normal distribution at every level. Therefore, non-parametric related samples

Wilcoxon ranked tests were used to compare within groups for each level of the test. The descriptive statistics and significance scores are shown in Table 6.14. The control group improved on level one and level three, while the experimental group improved on level two.

Table 6.14 Listening Vocabulary Levels Test results within group by level

	Experimental (<i>n</i> = 65)			Control (<i>n</i> = 51)		
LVL	<i>Mdn</i> score	<i>IQR (middle</i> <i>50% range)</i>	<i>Related-</i> <i>samples</i> <i>Wilcoxon</i> <i>ranked</i>	<i>Mdn</i> score	<i>IQR (middle</i> <i>50% range)</i>	<i>Related-</i> <i>samples</i> <i>Wilcoxon</i> <i>ranked</i>
Level 1 pre	23.0	2		23.0	2	
Level 1 post	23	1	.13	23.0	2	.007*
Level 2 pre	21	3		20	3	
Level 2 post	21	2.5	.010*	21	3	.204
Level 3 pre	16	5		15	5	
Level 3 post	17	4.5	.226	15	4	.012*

* $p < 0.05$

It is encouraging to see that both groups were able to improve their scores on the separate levels of this test, despite the words in the test not being taught or included in class materials. This analysis has shown a positive relationship between general vocabulary knowledge and ability to gain productive knowledge of multi-word expressions. The analysis also supports the finding from Study One and Study Two that the intervention has no comparable effect on general aural vocabulary knowledge. That brings us to the end of our pre- and post- measures and test results. What is left to consider is how the experimental participants (the classroom learners) felt about the classroom activities and which activities they considered the most useful for learning.

6.7 Which activities were considered most useful by the learners?

Participants from the experimental group completed online feedback surveys in class regarding their response to the activities. Participants were asked to respond showing the degree to which they agreed that each activity was useful for learning English or not. Participant response numbers for each survey and the percentage of participants who agreed that the activity was useful is shown in Table 6.15. The activities are listed in order from the activity receiving the most agreement that it was useful (by percentage).

The activity of recording the role-play had the highest percentage of agreement (88%) that it was a useful activity for Study Three. 88% agreement was an increase from 73% in Study One, and 76% in Study Two. It should be noted that in Study One there was a neutral response option which was not available for Study Two and Three, hence the percentages from Study One are likely to be lower due to this difference in survey response options. In this case, ‘useful’ did not equate to ‘enjoyable’, though, as only 53% of participants in Study Three considered recording role-play to be a fun activity.

Table 6.15 Activity usefulness as agreed by Study Three participants

Activity	<i>n</i>	% responded that activity was useful
Record role-play	<i>n</i> = 68	88.2%
Free role-play	<i>n</i> = 66	87.9%
Role-play	<i>n</i> = 72	87.5%
Phrase instruction	<i>n</i> = 67	86.6%
Decreasing time role-play	<i>n</i> = 66	84.8%
Shadowing	<i>n</i> = 62	82.3%
Mingle jigsaw	<i>n</i> = 67	74.6%
Dictogloss	<i>n</i> = 71	73.2%

Free role-play was considered useful by the second highest percentage of participants; with 88% considering it a useful activity for learning English. This rating was in contrast to Study One, where free role-play was considered useful by only 61%. For Study Two, scaffolding was added, and scenario choice removed, this adjustment resulted in the usefulness perception increasing to 80%. Unlike recording the role-play, free role-play was also considered fun by the majority (74%) in Study Three.

Simple role-play with notes has been consistently perceived as useful, in Study One by 73%, Study Two by 95% and Study Three 87.5%. During a focus group, further details were elicited. In the first two studies, volunteers had been exclusively female. However, despite there being only one focus group volunteer for Study Three, he was male. His pseudo name was Leo. He said, “of course role-play is useful, it is also fun trying to role-play without looking at the notes and speaking with different partners.”

In the focus group, Leo remarked that shadowing was difficult because most Japanese are trained to listen to slow spoken English and write marks between the words that they hear. This activity is done in middle school in order to recognize the

individual words in English. However, for shadowing, the audio speed is not artificially reduced, and the students must try and mimic this speed, leaving no extra time for recognizing and attaching meaning to each word.

Leo also mentioned feeling a sense of achievement or fulfilment when he was able to write and check his dictogloss because there were four opportunities to hear it. He also said that finishing the role-play within the time limit for the decreasing time role-play gave him a sense of fulfilment or satisfaction. This sense of achieving was what made those activities enjoyable for him. Certainly, giving students multiple opportunities to perform a task gives them the ability to notice where they can improve and then work on achieving it. Considering these results, it appears that all the experimental classroom activities were perceived as useful for studying English and most of the activities were considered enjoyable to some extent. The positive response from participants to the activities suggests that the activities are suitable for the classroom and can be recommended to teaching practitioners.

6.8 Study Three results summary

This chapter has detailed slight adjustments to the experimental classroom activities and presented results from Study Three. Results from this third study have also been compared with the previous two studies. The results from Study Three confirm through replication that focusing on multi-word expressions through classroom activities improves learner knowledge of form and meaning for the target multi-word expressions. Results also show an increase in use of the target multi-word expressions by participants in spoken dialogue after the intervention, a result that was not clear in the previous two studies. The increase in use of target multi-word expressions correlated positively with increase in speech rate and mean length of runs. The next chapter will discuss the main findings emerging from this research project as they relate to the wider literature and ongoing research.

7 Chapter Seven: Discussion

One of the most important issues that this research sought to address was whether EFL speaking fluency could improve through teaching and practice activities focused on multi-word expressions. The first study of three replicated a potential set of classroom activities for building knowledge and use of multi-word expressions. This approximate replication of Wood (2009) with more participants and a control group provides evidence that the fluency gains achieved in the original study through a focus on multi-word expressions cannot be easily generalised to a wider EFL population within a nine-hour intervention period. Using an action research model, participant feedback on the classroom experience was gathered and improvements made in response. This process revealed preferences of learners in the Japanese cultural context for structure and avoiding uncertainty. The second study improved upon the method of the first study in response to participant feedback. Having fewer participants provided the opportunity to delve deeper into individual learning. A closer investigation indicated the importance of familiarity with expressions for fluency improvement. The third study replicated the second and confirmed that classroom activities not only built up learner knowledge of the meaning and form of multi-word expressions, but also developed the students' ability to use multi-word expressions in conversation and confirmed the hypothesis that use of multi-word expressions correlates positively with fluency. However, nine hours were found to be insufficient in the EFL context to see an increase in spoken fluency when compared with a control group. Taken together, the three studies provide empirical evidence for the value of sequenced classroom activities which build up learner knowledge of target multi-word expressions and provide iterative experience using the expressions in interactive activities and conversations. Chapters Four through Six have investigated and provided answers to research questions surrounding the effectiveness of teaching multi-word expressions in the classroom. This chapter interprets the combined findings in light of relevant literature under four themes: (1) what we have learnt from replication and action research, (2) stages of skill development for chunking of target multi-word expressions, (3) how fluency develops in an EFL environment, and (4) teaching fluency as part of a language course.

7.1 Learning from replication of the fluency workshop

This research sought to confirm whether the increase in fluency based on teaching multi-word expressions reported in Wood (2009) could be extended to an EFL context. While Onoda (2014) had been inspired by the Wood (2009) fluency workshop model and had closely based interventions on the original activities in an EFL context, the absence of a control group limited interpretation of the results. More recently, McGuire & Larson-Hall (2018) also approximately replicated the original Wood (2009) study with a comparison group in an ESL environment, and they found significant fluency gains, which corroborates the results from the original study. The current study contributes to the field by approximately replicating the Wood (2009) study with a control group in an EFL context. This particular EFL context in northern Japan provided very few outside of class opportunities to use English, allowing knowledge and fluency gains to be almost entirely attributed to the intervention. Approximate replication allows investigation into the generalisability of the original Wood (2009) study. That is, can the results found in Wood (2009) be expected in EFL situations?

The approximate replication reported in Chapter Four revealed that while the intervention increased knowledge of multi-word expressions significantly more than the control condition, post-intervention use of multi-word expressions and fluency were not significantly different between the two groups. Taken alone the experimental group did make significant fluency gains, and so in this respect we can say that the conclusions of Wood (2009) and Onoda (2014) were supported since there was no control group in those studies. However, fluency gains were not significantly greater than the gains made by a control group, and therefore the fluency workshop intervention was statistically no better for building fluency than the control groups' linked skills course. The approximate replication showed that if a teacher/researcher were to use the class activities as in the original Wood (2009) fluency workshop intervention, they could expect to see fluency outcomes similar to that of a linked skills intervention in an EFL context over a nine-hour intervention. To some extent, the learning to be expected might change depending on the starting knowledge of the learner; a quick look at the results from Study Two reveals that some individuals did achieve great increases in fluency while others showed a reduction in fluency. Individual differences in motivation and

focus during the classes and tests would probably help to explain some of this variation. Unfortunately, I did not investigate learner motivation as a variable in this investigation. Overall, the three studies all revealed that one can expect to see an increase in productive form and meaning knowledge links for multi-word expressions taught and practiced. In order to probe how the experimental activities might be improved and perhaps thereby improve learning outcomes, I took an action research approach which I discuss in the next section.

7.1.1 How action research can inform activities in the classroom

Action research inquires how to improve classroom learning through gathering feedback from participants and researcher observations. Feedback was collected from participants in Study One and participant feedback combined with my own observations suggested that adaptations could be made to the Study One fluency workshop activities to improve their effectiveness and appropriateness for the context. Therefore, Study Two reported in Chapter Five was conducted which trialled changes and provided insight into the uptake of multi-word expressions and fluency gains at individual levels. Feedback from participants in Study Two when compared with the feedback from Study One suggested that the changes made to activities were effective and appropriate.

In the literature review, I introduced the possible cultural differences that might affect how appropriate activities are for the EFL context in Japan. The fluency workshop activities originated from Canada where individualism is strong and there is a certain acceptance of risk (Hofstede, 2011). In Japan, the opposite is true, group harmony is a stronger value than individualism and uncertainty avoidance is high. In each of the three studies, participants were surveyed regarding their feelings towards the classroom activities. Avoiding uncertainty in classroom activities was shown to be important for social comfort, and therefore activities were adapted to reduce feelings of uncertainty or social awkwardness. Adding time limits to activities such as the dictogloss and the mingle jigsaw, where participants had to initiate interaction with peers of their own accord appeared to help learners overcome to some extent this inclination to avoid uncertainty. The time limits helped participants focus more on achieving the task within the time frame, than on holding back from initiating

conversation due to social uncertainty (Kapp, 2012). Additional scaffolding and removing choices for the free role-play also appeared to help learners focus on speaking rather than worrying about making a pre-task choice between scenarios.

The adaptations appeared to be effective in Study Two, but it was not until Study Three, when there were more participants and a control condition, that we saw the use of multi-word expressions improve significantly. Participant responses to the Study Three activities also showed more satisfaction with the activities, with fewer suggestions for improvement. In particular, setting activities at a level where learners are challenged, but also have the ability to feel that they have achieved, was shown to be a motivating factor in Study Three. Creating achievable goals that learners can aim towards also links back to the gamification aspect of classroom activities (Kapp, 2012). When a learner experiences satisfaction by achieving a goal, they are positively motivated for further learning and achievement. Alone, any of these studies have limited insight, but when replicated and repeated with sensitivity to learner feedback and researcher observations, we can have more certainty about the results and an effective pattern for teaching in the Japanese EFL context emerges.

7.2 Stages of skill development for chunking multi-word expressions

An important outcome or message from this classroom-based study is sequencing of activities. Learning can be seen to occur in stages as the learner absorbs new concepts and seeks to enact them. It appears that the base for learning to use a multi-word expression comes from having built familiarity with the form and meaning of the expression and having exemplars of how it can be used in interaction. Jumping into use and practice without such a declarative knowledge base is like jumping into the deep end of a pool before learning how to float. The whole process flows more smoothly when done in sequential steps. Knowledge development is argued to develop in a linear three stage pattern according to the law of learning/practice (DeKeyser, 2015; Dörnyei, 2009) and skill acquisition theory beginning with 1) the cognitive/declarative/presentation stage, followed by the 2) associative/procedural/practice stage, and culminating in the 3) autonomous/automatic/production stage (Anderson, 2000; DeKeyser, 2015; Dörnyei,

2009). In this study I have used three measurements which line up nicely with the three stages of skill learning in regard to multi-word expressions; 1) the cognitive/declarative stage measured by ability to recall the form and meaning of multi-word expressions for a productive cloze test, 2) the associative/procedural stage measured by ability to use multi-word expressions in conversation, and 3) the autonomous/automatic stage measured by ability to use multi-word expressions in conversation with a fluency advantage. Results aligning to each of these skill acquisition stages will now be discussed in relation to relevant theory and literature.

7.2.1 Cognitive/declarative/presentation stage measured by cloze test

Experimental participants in all three studies were able to reliably demonstrate learning on the meaning-form link test (cloze test) for the target multi-word expressions. This evidence that declarative knowledge develops robustly for the multi-word expressions taught in the intervention provides insight that previous research (McGuire & Larson-Hall, 2018; Onoda, 2014; Wood, 2009) has not touched upon. The improvement was not due to any testing effect since the control group took the same tests. The strong results in all three studies can be interpreted through the Anderson (1983, 2000) model of cognition, whereby adult learner attention is first focused on developing cognitive or declarative knowledge (encoding the meaning and form of language items). My hypothesis was that the participants who received clear presentation of the meaning and use of the multi-word expressions combined with multiple encounters and practice opportunities would improve their knowledge of meaning and form for these expressions. This hypothesis held true with the results confirming that the focus on multi-word expressions was effective for developing ability to retrieve and produce meaning and form knowledge of the expressions when prompted. Learning from the intervention appeared to be quite successful in that this cloze test was a productive test. Receptive knowledge is easier to gain and demonstrate than productive knowledge (Webb & Kagimoto, 2009). Receptive knowledge could have been tested by presenting the L2 and asking participants to select the meaning from multi-choice options, for example see Schmitt et al. (2004). Such a receptive measure would likely show the greatest amount of learning, as receptive knowledge is known to be a pre-requisite to productive knowledge. Significant gains in the productive

cloze tests are testament to the effectiveness of the intervention for developing productive knowledge of multi-word expressions. In the next section, I will discuss what this study has revealed regarding the connection between chunking and vocabulary knowledge.

7.2.2 Chunking and connection with vocabulary knowledge

Chunking referred to earlier in Chapter Two, is a technical term in psychology which refers to our natural strategy of associating smaller units into larger units thereby freeing up limited short-term memory capacity (Bybee, 2008; Dörnyei, 2009; N. C. Ellis, 2001). Chunking theory states that the more chunks a learner knows, the greater their ability to create “higher level chunks,” which extend from what they already know (N. C. Ellis, 2001, p. 39). In SLA, existing knowledge of individual words such as; *I*, *will*, and *think* is likely to assist learning and association of those individual words into the higher order chunk ‘*I think I will.*’ Therefore, the learner who already knows the individual words that make up a chunk has an advantage over the learner who does not know those words. My hypothesis therefore was that learners with greater vocabulary knowledge would be able to learn the 4-grams (multi-word expressions) faster because they likely had greater knowledge of the words or bigrams that made up the expressions and would therefore be able to make faster associations between words they knew into higher level chunks (N. C. Ellis, 2001; Kasahara, 2011; X. Zhang, 2017). There were positive correlations between vocabulary knowledge (measured by the pre-intervention Listening Vocabulary Levels Test) and improvement in multi-word expression knowledge after the intervention in Study Three, which supports this hypothesis. Quantitative evidence from this study suggests that learners with greater vocabulary knowledge will be more efficient in general at learning multi-word expressions than those with less vocabulary knowledge.

Qualitative evidence also suggested that existing knowledge of component words of a multi-word expression presents an advantage for learning the multi-word expression and achieving a greater fluency. In Study Two, there were only two instances where complete expressions were used. The use of the multi-word expression had different fluency outcomes. In the first case, the speaker showed no knowledge of

the expression in the pre-test but managed to use it correctly in the post-test conversation accompanied by a decrease in fluency. The decrease in fluency suggests that while the speaker could use the expression, it was an effort for their working memory to retrieve it and use it. In the second case, a different speaker showed partial familiarity with an expression through the pre-intervention cloze test. This speaker was able to use that same expression correctly in the post-test combined with an increase in fluency, providing supporting evidence that greater knowledge of component parts leads to faster automatization of the whole expression.

7.2.3 Associative/procedural stage (multi-word expression use)

Cognitive or declarative knowledge is widely accepted as the necessary precursor to the development of procedural knowledge and automatic knowledge (Anderson, 2000; DeKeyser, 2015; Dörnyei, 2009). The hypothesis in this study therefore was that learners would show an increase in meaning and form recall knowledge before showing an increase in unprompted use of multi-word expressions in spontaneous speech. The multi-word expression cloze test results confirmed this hypothesis showing that the first stage of declarative learning had indeed taken place ahead of procedural knowledge development as seen in Study One and Two. It was clear that the participants in Study One and Two had improved their declarative knowledge of the expressions but not their procedural knowledge. The lack of use (despite knowledge) of the expressions indicates that learners are only willing to use expressions in conversation after they are sufficiently confident with the meaning and form links. Avoidance of areas where error could occur might indicate a focus on accuracy, but could equally be suggestive of avoidance of uncertainty (Skehan & Foster, 2001). Both of these explanations are plausible for the Japanese learner stereotype.

7.2.4 Number of encounters

As the number of encounters with language items increases, so does familiarity, confidence and the ability to produce the items in conversation (Bybee & Hopper, 2001; Joe, 1998; Uchihara, Webb, & Yanagisawa, 2019). My hypothesis was that increasing the number of encounters with the expressions would increase the associative or procedural knowledge of the expressions (ability to use the expressions in

conversation). In order to increase encounters with the expressions, the classroom activities needed to be optimised. Optimisation occurred in Study Two and Study Three through adding time limits to activities, displaying the expressions more often, and reducing cognitive load for surrounding details through increasing scaffolding. Time limits are known to motivate action and focus on the task, contributing to game like conditions (Kapp, 2012). Adding time limits to communicative activities such as dictogloss and mingle jigsaw visibly helped learners to overcome inhibitions and put effort into achieving the activity within the timeframe. The addition of the time limit therefore likely increased the number of encounters with target expressions, as more efficient use of time would probably result in more expressions able to be used per minute. Also, increased display of the target expressions on the class screen during activities such as the decreasing time role-play and free role-play, where notes were not allowed, increased the opportunities to read and use the expressions. Display of the expressions would have been particularly beneficial for those newly learned expressions not yet known well enough to retrieve from memory for use. In the third study, realia scaffolding for hotel booking forms were also provided which helped participants remember the information they needed to ask about when role-playing, serving again to increase opportunities to use expressions as less time was spent wondering what information had not yet been asked. The improvements made to the activities between Study One and Study Three contributed to and increased encounters with the multi-word expressions which in turn led to an increase in associative or procedural use of target expressions in the post-measure conversations in Study Three.

In line with the number of encounters principle, an overlap of some word sets between the expressions also seemed to increase use of those word sequences (bigrams and trigrams) in conversation in Study Three. These findings support the hypothesis that the more encounters learners have with multi-word expressions, the more likely they are to use the expressions in conversation. The learning from massed repetition may actually be more able to withstand *the forgetting curve* (Ebbinghaus, 1964) than exposure to various vocabulary items that are spread out more over time. This study did not attempt to compare massed repetition (one day/session of exposure) with spaced repetition (exposure over several days/sessions), but in so far that use of multi-word

expressions was concentrated during the activities, the results from this study agree with the meta-analysis by Uchihara et al. (2019) whereby massed repetition showed stronger correlations with word learning than spaced repetition.

7.3 Automatic stage: Relationship between multi-word expression use and fluency

An important hypothesis motivating this study based on current literature (Boers et al., 2006; Kuiper, 2004; McGuire & Larson-Hall, 2018; Wood, 2009) was that speaking fluency would improve as more multi-word expressions were used. Fluency (referred to here as the autonomous or automatic stage) is the third stage of skill acquisition. I expected to see a positive correlation between use of multi-word expressions and measures of fluency (speech rate, phonation time ratio, articulation rate, and mean length of runs). The experimental group in Study Three was the only group that showed a significant increase in spoken use of the target expressions compared with a control group, so I will discuss the correlations between fluency measures and use of multi-word expressions that were exhibited in Study Three. The experimental group showed within group increases in fluency measured by speech rate. Speech rate showed a moderate positive correlation with multi-word expression use $rs(46) = .39, p = .008$. Phonation time ratio (% of time speaking including pause time) also significantly increased within group and was mildly correlated with multi-word expression use $rs(46) = .32, p = .03$. Therefore, speech rate and phonation time ratio results support the hypothesis that speaking fluency improves with use of multi-word expressions.

However, the articulation rate (essentially the speech rate without pauses) did not improve in Study Three. The correlation between articulation rate and use of multi-word expressions was positive though weak; $rs(46) = .24, p = .11$. N. de Jong & Perfetti (2011, p. 539) argue that the articulation rate “measures the speed of articulatory processes” without having a strong relationship with proceduralisation of lexis or syntax. Articulation rate increases when expressions are spoken quickly, particularly with phonetic reduction. In the case of L1 speakers, phonetic reduction often occurs for “highly probable content and function words...in conversational speech”, while for L2 speakers phonetic reduction is linked to familiarity with the expression (Siyanova-Chanturia & Van Lanker Sidtis, 2019, p. 49). These results indicate that the speakers

were not familiar enough with the expressions to speak them at speed or with phonetic reduction. In Study Three, I interpreted the unchanged articulation rate, despite a reduction in pause time, as indicative of slow speaking and a strain on working memory to produce expressions. It is also possible that self-monitoring was slowing articulation.

Like articulation rate, there was also no significant change in mean length of runs (syllables uttered between pauses of .25 seconds) for either group. Towell et al., (1996) argue that proceduralisation of language can be seen by an increase in mean length of runs when the phonation time ratio is stable. The increase in use of multi-word expressions, coupled with no increase in mean length of runs, reveals that the multi-word expressions were not known well enough to be used without internal pausing. Prosodic coherence was not a criterion for counting multi-word expression use in the transcripts, therefore it was possible for the use of multi-word expressions to increase without a parallel increase in the mean length of runs. Despite not increasing, mean length of runs was positively correlated with use of multi-word expressions $r_{s(46)} = .51$, $p < .001$, which supports the original hypothesis that use of multi-word expressions is associated with fluency.

The correlations between use of multi-word expressions and the four fluency measures were all positive. Mean length of runs showing the strongest correlation, followed by speech rate, phonation time ratio and the weakest correlation was with articulation rate. Therefore, this study supports the hypothesis that fluency improves as use of multi-word expressions increases.

7.3.1 Building fluency in an EFL context: Nine hours is not enough

The hypothesis that the intervention would result in greater fluency for the experimental group over the control group was based on the literature claiming that use of multi-word expressions is associated with greater fluency (Boers et al., 2006; Kuiper, 2004; McGuire & Larson-Hall, 2018; Wood, 2009). While the correlations were positive, the difference in fluency between the experimental and control groups was not significant, which seems contradictory. It is important to point out that the commonly referred to literature which associates fluency with multi-word expressions is not based on nine hours in an EFL context. Kuiper's (2004) findings were based on native

speakers rather than language learners. Boers et.al. (2006) was based on 22 intervention hours in the EFL context of Belgium; much longer than the nine hours of this intervention. Meanwhile, the studies by McGuire & Larson-Hall (2018); and Wood (2009) were conducted in ESL contexts, with more opportunities than the EFL context for natural input and practice. Other studies based in Japanese EFL classrooms investigating speaking fluency have found fluency increases over one academic year, though without control groups (Herder & Sholdt, 2014; Onoda, 2014). By comparison, in this project we saw a significant increase within the experimental group in speech rate and phonation time ratio over only nine intervention hours relative to the beginning of the course. Considering the timeframes of the above studies, this intervention was by comparison fairly successful for building fluency.

Everything considered, the participants in my study did not have a firm enough grasp on the target expressions to use them fluently in conversation. In Study Three, the experimental participants were displaying the first two stages in skill acquisition theory. The cognitive/declarative stage was achieved and seen through strong results in the cloze test, and the associative/procedural stage was achieved and seen through increased use of the expressions in conversation in Study Three. All the same, the automatization/automatic stage which Dörnyei (2009) considers as ‘fluency’ was not reached, and the fluency improvements were not greater than the control group. Given more practice and time, we would expect all measures of fluency to improve based on the power law of practice/learning (DeKeyser, 2015; Dörnyei, 2009) that sees reduced reaction time and greater accuracy developing steeply with practise until a plateau is reached where use is considered fluent or automatic. Reaching that automatization stage takes longer than the time to develop the first two skill levels and “language classrooms are usually unable to provide the learner with the large amount of practice time that is needed for the completion of the automatization process” (Dörnyei, 2009, p. 290).

Let us recall the three frequency effects (Bybee, 2014) described in Chapter Two, the first was the conserving effect whereby repetition leads to easier retrieval. It was evident that learners were able to retrieve and use the expressions more which suggests the conserving effect was active. The second frequency effect expected was autonomy, whereby the expression stands alone rather than being constructed word by word. In this

study, I did not measure phonological coherence of expressions specifically, so no evidence to this effect was presented. The third frequency effect was the reducing effect, whereby phonological reduction occurs as the expression is quickly spoken. This reducing effect was not seen in the present study, as evidenced by the lack of change in the articulation rate. Therefore, of the three frequency effects, only one was clearly achieved; the conserving effect.

It seems that the reducing effect would not appear without much more intervention to increase familiarity with the expressions in the EFL context. Certainly, the starting language proficiency, familiarity with the multi-word expressions, and familiarity with the classroom activities will influence uptake and resulting fluency gains. Individual differences in uptake cannot be understated. For low proficiency learners with little familiarity with the multi-word expressions, it seems unlikely that a significant increase in speaking fluency over nine hours in an EFL context will occur, particularly when there are limited opportunities to use English outside of the class. The results of this study could be interpreted as revealing an in-between acquisition zone whereby learners have linked knowledge of form and meaning with the target expressions and are able to use them with conscious effort but are still developing automaticity/fluency for using the multi-word expressions effortlessly in conversation. The next section discusses what this research has revealed about teaching fluency as part of a language course.

7.4 Four Strands and linked skills in a language course

There was a common foundation that both the control and experimental courses were based on: the Four Strands principle of balancing meaning focused input, language focused learning, meaning focused output and fluency activities (Nation & Yamamoto, 2012). Both courses could also be described as linked skills based. It is important to note that the sequence of the strands in this study started with declarative knowledge building through meaning focused input and language focussed learning before progressing to practice/developing procedural knowledge through meaning focused output and finally fluency activities in order to build automaticity/fluency. A good balance of the Four Strands is argued to create the best conditions for fluency practice.

While the balance of the Four Strands was not strictly adhered to, the linked skills base meant that all Four Strands were addressed in both courses. Despite the experimental group spending a greater proportion of course time doing fluency activities, their fluency gains were not significantly greater than the control group, suggestive that a greater focus on fluency activities may not necessarily lead to greater fluency when compared with a sequenced linked skills program that covered all Four Strands. The control condition did not focus on the target expressions or on speaking fluency per se, but they had a linked skills syllabus which developed their experience with the language needed for their course (Nation & Newton, 2009). The control group and the experimental group differed in the variety of input materials, with the control group arguably having a greater range of input and output genres and vocabulary. In contrast, the experimental group used one model dialogue for all of their class input and output activities over two class sessions. The experimental group were repeating a model dialogue through the same writing and speaking activities for each of the three units. The sequenced build-up in both conditions to fluency activities may help to explain to some extent the lack of difference between the groups in terms of fluency development. The outcomes over a six-week period in an EFL context suggest that a linked skills course such as the control group had does not differ significantly for building speaking fluency from a course focussed on multi-word expressions and speaking fluency.

I wondered whether the control group might improve their general vocabulary learning more than the experimental group because they were encountering a wider range of vocabulary and input texts. In Study Three, the difference between the groups for change in aural vocabulary was approaching significance with a small effect size. Despite the experimental group having a narrowly concentrated and repetitive program compared with the control group, there was no significant difference between the two groups for overall general vocabulary learning. This result suggests that over nine hours in an EFL context, neither a concentration on multi-word expressions nor a wider ranging linked skills course shows any advantage over the other for general vocabulary learning.

7.5 Conclusion

In this chapter, I have discussed how the findings from this research project can be interpreted through and contribute to the wider SLA literature. In particular, this project has shown the value of replication for expanding our understanding of how the teaching of multi-word expressions affects knowledge and fluency outcomes in an EFL context. The action research model of gathering and responding to feedback from participants was shown to be valuable for improving the effectiveness of classroom-based learning. Insights gained through learner feedback enabled learning activities to be adapted to take advantage of cultural preferences in order to improve the classroom learning experience for both learners and teacher.

Focussed teaching and practice of transparent multi-word expressions over nine hours in an EFL context was shown to result in clear gains in multi-word expression knowledge at the cognitive/declarative and associative/procedural skill level. The hypothesis that skill acquisition occurs in linear progression from declarative to procedural to automatic was supported. The experimental activities (listening for gist, shadowing, mingle jigsaw, dictogloss, role-play, decreasing time role-play, recording and reflecting on performance and free role-play) were shown to lead to both declarative and procedural knowledge of target multi-word expressions.

Results supported the hypothesis that multi-word expression use was positively correlated with fluency measures. Based on the literature and the results showing development from declarative to procedural skill, it seems likely that more practice with the multi-word expressions would result in automation or fluency improvement. However, whether further class time should be used to attempt to gain greater fluency using the targeted multi-word expressions is debateable and would depend on the learner goals. Extending the class time spent on specific multi-word expressions using the same intervention risks boredom on the part of learners and would also represent an opportunity cost for missed learning of other language features. Such a narrow focus would need to be carefully considered. Rather, I would recommend deliberately re-meeting multi-word expressions in future units (without specific attention) to help learners recall and consolidate their learning. Finally, Nation's (2007) Four Strands structure which was the underlying base for both conditions was found to be effective in

supporting fluency development. The implications of this research for teaching, along with the limitations of the research and future research directions will be addressed in Chapter Eight.

8 Chapter Eight: Implications and conclusion

As a classroom teacher, I was motivated to start this research in the search for classroom interventions to help EFL learners improve their speaking fluency. While the literature suggests that using multi-word expressions increases speaking fluency, evidence regarding how to effectively teach multi-word expressions so that learners can use them in their conversation has been limited. Therefore, the current research fills some of the gaps regarding how to teach multi-word expressions to language learners in the classroom, specifically in the EFL context in Japan. In this chapter, I will address the summative claims arising from the research project with theoretical implications. Then I will describe the pedagogical implications for classroom learning contexts. Following on from that, I will reflect on the limitations of this research and some emerging methodological implications. There were some research questions that I was able to answer through this research while some new research questions have emerged; therefore, I will shine a light towards future research directions. Finally, I will conclude the chapter (and thesis) with my reflections on the PhD journey.

8.1 Summative claims and theoretical implications

In this section I will make summative claims based on the results in regards to the positive relationship between use of multi-word expressions and fluency, vocabulary knowledge or proficiency and uptake of multi-word expressions, and support for both the skill acquisition theory (DeKeyser, 2015) and the number of repetitions or encounters positively correlating with learning of the item (Joe, 1998; Saragi et al., 1978; Uchihara et al., 2019).

Firstly, this study found that the use of multi-word expressions appeared to enhance learner fluency. A positive correlation was evident in Study Three between use of multi-word expressions and four measures of fluency; speech rate, phonation time ratio, articulation rate and mean length of runs. This outcome corroborates previous research results which have associated greater use of multi-word expressions with increased fluency (Boers et al., 2006; Kuiper, 2004; McGuire & Larson-Hall, 2018; Wood, 2009). The positive relationship between use of multi-word expressions and

fluency heightens the importance of helping learners to use such multi-word expressions in conversation.

This research has also revealed a positive relationship between general vocabulary knowledge and learning speed for multi-word expressions. This relationship suggests that those with greater vocabulary knowledge can be expected to learn multi-word expressions at a faster rate. This theory extends the chunking theory (N. C. Ellis, 2001) whereby learners build associations between the words they already know. The extension to the theory supported by the results of this research is that if learners already know many words, they can be expected to make associations between words with less effort and therefore achieve faster acquisition. Vocabulary knowledge was also shown to correlate with proficiency in this investigation, which supports previous research (McLean et al., 2015; Stæhr, 2008). By extension, one can also theorise that greater starting language proficiency results in more efficient learning of multi-word expressions.

The results from this research also support the cognitive/skill acquisition theory (Anderson, 1983, 2000; DeKeyser, 2015); that declarative knowledge development precedes procedural and automated knowledge development. We saw productive recall of meaning and form for the multi-word expressions developing faster than the ability to use multi-word expressions in conversation for EFL learners. Declarative knowledge was evidenced in the productive recall cloze test. Subsequent procedural knowledge (ability to use) the expressions was seen in increased use in post-intervention dialogues in Study Three. Increases in automation (spoken fluency) were also seen through medium effect sizes within the experimental groups in all three studies. Automaticity was however the only area of skill acquisition where the experimental group were unable to achieve significantly greater gains than the control group. The progression from declarative, procedural to automatic skill appears to explain skill development for learning multi-word expressions in this research.

This research has shown furthermore that the more often a multi-word expression is encountered, the more likely it is that the speakers will be able to use the expression in conversation. Layering or massing multiple practice opportunities using the

expressions seems to increase ability to use multi-word expressions. This finding supports the theory that the more encounters had with a language item, the better it is learnt (Bybee & Hopper, 2001; Joe, 1998; Uchihara et al., 2019). In summary, the theoretical implications of this research are; that a positive relationship can be expected between use of multi-word expressions and fluency measures, that existing vocabulary knowledge and proficiency supports greater uptake of multi-word expressions, that we can expect declarative knowledge to precede and be followed by procedural and automatic knowledge, and that increased encounters lead to greater ability to use multi-word expressions. In the next section, I will highlight the pedagogical implications emerging from the research.

8.2 Pedagogical implications

This research was motivated from the very beginning to improve classroom teaching and learning. It is therefore fitting to share the many pedagogical implications that this research offers to the language teaching field. Perhaps the most pressing question for practitioners is whether to give multi-word expressions special attention, given that every class has different goals. My recommendation based on this research experience is that if building familiarity with common expressions is a goal then teachers will definitely want to give some focussed class time to multi-word expressions. In this section, I will suggest how to select multi-word expressions for learning, using context and frequency which were my guiding selection methods. I will then recommend activities used in this study that highlight, teach and provide interactive practice for using multi-word expressions in the classroom. I will then describe how action research can be used by teachers to discover more about their learners and their learning preferences. Finally, I will end this section by discussing the cultural preference for structure that I found among my Japanese EFL learners, and I will recommend adaptations that can be made to take advantage of such learning preferences.

8.2.1 Selection of multi-word expressions

In my study, I identified multi-word expressions based on contextual use and frequency profile. Teaching practitioners (inclusive of teachers, program designers,

materials and textbook writers) need to identify multi-word expressions, perhaps even in current classroom materials, that they consider useful for learning. To decide which expressions are useful, we need to consider which situations learners desire to use the language to communicate in. A survey of learner goals could provide valuable perspectives here. It is a logical first step to focus on preparing for learner anticipated speaking situations. The themes used in the experimental condition imagined using English to communicate on the street, in hotels and cafés. These were all possible future scenarios for the students identified by a previous cohort and hence the learning value was clear.

After identifying possible multi-word expressions for attention, the frequency profile of the potential multi-word expressions should be checked with appropriate corpora. The appearance or not of a multi-word expression in a corpus can depend a lot on what type of texts the corpus is made up of. Therefore, if conversational interactions are the goal, spoken corpora should be used to check the multi-word expression frequency profile. Currently there is a spoken sub-corpus from the Corpus of Contemporary American English (COCA) (Davies, 2013) that can be searched for *n*-gram (multi-word expression) frequency of occurrence. The multi-word expressions used in this project were checked using that spoken sub-corpus. The next section will describe how the multi-word expressions can be noticed and practised using the Four Strands; meaning focused input, language focused learning, meaning focused output and fluency activities.

8.2.2 Activity sequencing for fluency

One key pedagogical application of this research lies in the sequencing of activities for fluency development in classroom settings. In my study, based on the Wood (2009) fluency workshop, meaning focused input, explicit noticing and language learning activities were provided at the beginning of the activity sequence to notice the multi-word expressions in context and explicitly learn their form, pronunciation and meaning. Next, repeated output opportunities were provided (Nation & Newton, 2009). Repeated use of the multi-word expressions was an important feature of the activity sequence (Uchihara et al., 2019). Massed retrieval opportunities (repeated opportunities

within one session) and spaced retrieval opportunities (with temporal gaps between use and spread over more than one learning session) were naturally provided over the class sessions which likely enhanced procedural learning (Ullman & Lovelett, 2018). For example, a multi-word expression would be retrieved to use in a role-play and several minutes later retrieved again to use in another role-play. Each unit with 10 multi-word expressions was spaced over two class sessions, so there was also the space of a week between further retrieval of the multi-word expressions. The learning sequence built up to fluency activities, where pressure to perform at an enhanced speed and levels were provided.

The sequencing of the activities are an effective example of linked skills and Nation's Four Strands (Nation, 2014), though one could argue that the meaning focused output and fluency activities were allocated more time than the language focused learning activities. Therefore, I recommend that teachers use the fluency workshop model with their own classroom texts relevant to student needs; and re-use a selected text in layered activities such as those shown in this study. Teachers should direct learners to notice embedded multi-word expressions through explicit teaching of the meaning of multi-word expressions and provide pronunciation practice. Learners should then be given opportunities to re-use the multi-word expressions in context through activities such as shadowing, mingle jigsaw, dictogloss, and role-play with multiple partners, and the 4/3/2 increasing time pressure role-play. Also, activities done without notes such as 4/3/2 are best supported by displaying multi-word expressions so that learners were reminded to use them and can refer to them when needed. Displaying the multi-word expressions, even when notes are not to be used, enables further input and output opportunities.

In different contexts, I recommend trialling and adapting the fluency workshop activities and then removing the activities that students are less receptive to, or the activities that seem less effective. Such an approach would be one way to shorten or simplify the full activity sequence. For lower proficiency learners, I would recommend removing the free role-play activity. Nevertheless, with higher proficiency learners the free role-play could be really beneficial. The activities are likely to be more or less

useful depending on the context and the proficiency of learners. In the next section, I will make recommendations for implementing action research in the classroom.

8.2.3 Action research and learner feedback for improved classroom learning

As a teacher, I found the iterative action research model very informative for improving teaching practice. Action research is where a teacher seeks feedback on the classroom experience and makes improvements in response (Burns, 2010). The fluency workshop activity set was reported by Wood (2009) to increase speaking fluency for a Japanese ESL learner in a Canadian homestay context. However, I was teaching within a Japanese EFL context, so given the difference in learning context and learners, it was likely I would need to make adaptations to the fluency workshop activities.

It is important to find out how activities are perceived by learners. I sought feedback from learners in each study in the form of activity feedback surveys and focus group interviews. In each of the results sections of this thesis, I have reported and interpreted learner feedback. With each study for this project I learnt more about how the students responded to the various classroom activities. Observing their response and collecting their feedback, I was able to make adjustments for the next iteration in order to improve the effectiveness of the classroom activities. The improved results in the later studies can be traced back to responding to feedback and observations with appropriate adaptations to the classroom activities.

Creating space to collect learner feedback and further improve learning activities in response was perhaps one of the most important aspects of this research project. Giving learners a voice in the classroom was key for being able to understand the effectiveness or ineffectiveness of classroom activities. Asking students how they feel about classroom activities can be very informative. I had a different cultural background to my students, which meant that there was much that I could wrongly assume. I was from New Zealand and my students were mostly from Japan, with a few also from China. For example, my first impression of the shadowing activity was that students would be embarrassed to mimic the audio not just once but four times. However, participants told me through feedback surveys that shadowing was one of their preferred activities in Study One. They even responded that they would be happy to shadow in

unison as a class without embarrassment. English pronunciation is quite different from Japanese, so articulation practice is valued by Japanese learners of English. This appreciation of shadowing is worth noting for any other learning context where L1 pronunciation is quite different to the target language. Had I not explicitly asked for feedback on this activity, I would not have known their preference and may have even been tempted to remove the activity, since it was not a communicative activity. My observation was naturally one sided and interpreted through my cultural lenses, so student feedback provided insight that I may not have understood otherwise. The learners also demonstrated through feedback that they enjoyed the classes overall and even found some of the activities to be fun; this kind of feedback was also useful to me.

Some of the important themes revealed through the voices of learners in this project were that activities can be difficult not only for linguistic reasons but also because of social and technical reasons. While some linguistic difficulty is perhaps to be expected in the language classroom, social and technical difficulty can be an undesirable distraction from language learning. As teachers, I propose that we need to reduce social and technical barriers to using the target language in the classroom as much as possible because valuable class time can be easily lost when such difficulties are experienced.

Social difficulty was experienced when activities lacked clear structure or roles. Participants considered activities such as the mingle jigsaw more difficult because they were free to choose who they spoke to. In Study One, the mingle jigsaw was reported as being socially difficult by participants “less English and more communication ability,” “I can ask friends, but with others I have to consider whether they are open to it.” One way that social inertia can be mitigated is through setting challenging time limits on interactive activities. For example, introducing time limits during the mingle jigsaw activity from Study Two onwards mitigated some of the hesitation for social reasons that learners felt in Study One. Feedback from Study Two participants showed an improvement in the enjoyment of this activity. Every cultural context is different, but with learners in Japan I recommend adding time pressure for socially interactive activities such as mingle jigsaw and dictogloss to encourage efficient use of class time for learning.

Another way to mitigate social difficulty could be to assign partners rather than asking students to choose their own partners or groups and assigning topics rather than asking students to choose their own. For example, the perception of usefulness and enjoyment for the free role-play improved after removing the scenario options and providing greater scaffolding, which served to reduce the amount of uncertainty involved in the activity.

Realising that wider cultural values can affect behaviour in the classroom suggests that wider cultural preferences may also be relevant in other teaching contexts. According to Hofstede's cultural differences research (2001), Japan has high uncertainty avoidance tendencies. The learning preferences I found are therefore likely to extend beyond my sample and into the wider Japanese population and possibly to other countries with high uncertainty avoidance indexes. Learners in Japan and in other countries with high uncertainty avoidance indexes are likely to respond better to classroom activities that have a clear flow and structure. The challenge is striking a balance between the security of structure and the risk that accompanies free production. Students need to feel secure enough to take risks to use new expressions in their conversation.

Learners also referred to technical difficulty in their feedback, such as struggling to follow instructions for recording and uploading a conversation to the class Moodle page in the recording role-play activity. Technical difficulties involved with class activities can quickly use up class time leaving less time for language practice and learning. Activities which require such technical time investment from learners should be identified in planning. Time required for technical aspects of class activities should be minimised through careful planning. Any activity that requires noticeable time on technical aspects should have a clear learning outcome which justifies the time spent. If I was not doing research, I would reduce the time and complexity involved in this activity by removing the requirement to upload the conversation. The learners could simply listen back to the recording from their device and make reflections on their fluency based on the recording.

Borg (2010, p. 422) states that "language teacher research engagement is a minority activity", but it does not have to be. The action research style that I used for this project is an accessible research paradigm that teachers can use to make gradual

improvements to their teaching practices. A simple way to start action research is to survey learners after an activity asking what they appreciated about the activity and how they think the activity could be improved. After collecting learner opinions, the teacher can then reflect upon the feedback along with their own observations and consider how the activity could be improved. Adaptions can then be trialled, and learners surveyed again. Creating a simple feedback loop like this will help teachers to adapt classroom activities to their context, culture and learner preferences. In the next section I will consider the limitations of this research and the implications for future methodology.

8.3 Limitations and implications for methodology

This study was not without limitations. In this section, I will draw attention to the limitations from most important to least. Limitations in the following areas will be addressed: the six-week time frame, difficulties gathering data from learners, comparability between research data sets, and the difficulty of measuring procedural knowledge.

8.3.1 Six-week time frame

The overarching research questions centred around whether fluency gains would be possible after an intervention in an EFL environment. The intervention was modelled on the Wood (2009) study which was originally set in an ESL immersion environment. The Wood (2009) model had a six-week time frame which translated into nine hours of classroom-based intervention in the EFL context, and this short intervention time turned out to be a limitation. In the Wood (2009) ESL context, there would have been more interaction using English outside of the classroom. However, in the EFL environment in this study, interaction in English outside of the classroom was highly unlikely. There was a significant increase in fluency (measured by untrimmed speech rate) within the experimental group, but the increase was not significant when compared with the control group. The strong results within the experimental group suggest that given more intervention time, the fluency gains for the experimental group may have become stronger than the control group. Therefore, for future fluency intervention studies in EFL contexts, I would recommend classroom intervention time of more than 10 hours.

8.3.2 Difficulties gathering data from learners

The results from the spoken dialogues (measurements of speaking fluency and use of multi-word expressions in conversation) were slightly underpowered because there were fewer than 30 participants in each group. The lack of participant numbers stemmed from problems with data gathering. For example, when listening back to the pre- and post-intervention dialogues, I struggled at times to identify who was speaking due to similarity in voice tone and/or quiet voice samples. I asked participants to play the same role that they had played in the pre-intervention recording when they were doing the post-intervention recording, but unfortunately about half played the opposite role in the post-intervention recording. I decided not to use data that could not be reliably attributed to a speaker, or conversations where participants played opposite roles to their pre-intervention recording. Losing data in this way meant that there were fewer samples in the data set. For future research, where participant voices are likely to sound similar, I would recommend video recording to make it easier to identify who is speaking. I did not use video for this study due to the massive data storage issue it would have created with so many recordings. I would also recommend showing participants their role before the post-intervention recording to ensure that they would know to play that same role for the post-intervention dialogues.

One factor influencing fluency in the post-intervention conversation could have been that the control group likely saw little connection between the post-intervention role-play and their class content, so they would have simply participated in the post-intervention role-play as a fun activity using whatever words they liked to complete it. However, the experimental group were likely to see the connections between what they had studied and the post-intervention role-play, and therefore it is possible that they felt pressure to retrieve and use the phrases they had learnt even when it took them longer to do so. Such awareness may have reduced fluency but at the same time may have increased use of multi-word expressions for the experimental group in the post-intervention role-play. In the next section I will consider the comparability between the research data sets.

8.3.3 Comparability between research data sets

One advantage of repeating research is that earlier iterations can be compared with later studies. My ability to make direct comparisons between Study One and Study Three was limited because I made three changes between the studies. First, I changed the role-play instructions from Study One to Study Two. This change was aimed at increasing the range of multi-word expressions that participants might consider using, however the new instructions were also more complex. The extra effort exerted to understand and follow the instructions may have inhibited fluent conversation, more so than the original instructions. Second, task familiarity changed between Study One and Study Three for the control group. In Study One, the control group had no familiarity with the recording task whereas both groups in Study Three were given experience with a recording task before the study commenced. Third, counting of multi-word expression use in Study Two and Study Three was more inclusive of variation than in Study One. The counting criteria in Study One was strict but extended multi-word expressions were counted to encompass variation. However, in Study Two and Study Three the multi-word expressions use count encompassed some variation that was not included in Study One, but the count did not extend to the extent of variation found in the Study One extended multi-word expressions. Therefore, when making a direct comparison between Study One multi-word expression use and the later studies, this difference needs to be considered. These changes to role-play instructions, task familiarity, and multi-word expression counting method make direct comparison across the studies complicated because the cause of differences between studies could be associated with the change in instructions, task familiarity and/or multi-word expression use counting method rather than factors of interest such as increased encounters with the expressions. If I were to replicate this research, I would use the same role-play instructions for all iterative studies, make sure all participants were familiarised with the recording task and keep the counting method consistent with that used in Study Two and Three.

8.3.4 Difficulty measuring procedural knowledge

Measuring procedural knowledge (use of multi-word expressions) was limited for two reasons. First, the measurement was indirect and second, the measurements could

have been checked more thoroughly. Measuring ability to use multi-word expressions through dialogue is an indirect measurement of procedural knowledge. There was no necessity (task-essentialness) or prompting to use the expressions in the dialogue (Thomson et al., 2019) whereas there was in the cloze test. For example, participants could complete the role-play without using the multi-word expressions that had been practised, whereas in the cloze test they needed to produce the multi-word expressions to complete the test. This consideration means that gains in procedural knowledge may have been greater than shown through the dialogue measure.

For counting multi-word expression use in the transcripts, I thought that there would be no need for inter-rater reliability because I had a list of 30 expressions, so the subjective judgement needed to decide whether or not to count an expression was limited. However, on reflection there was a certain amount of subjective judgement used when deciding whether to count or not count variation to the target multi-word expression, so it would have been prudent to have another person check at least 10% of the multi-word expression use counting.

Also, while the timestamping for fluency measures was triple checked, it would have been ideal to have an inter-rater check over the accuracy of the timestamping for at least 10% of the data set. The PRAAT time stamping of pauses and speaking turns for the pre- and post-intervention dialogue recordings was first time-stamped and labelled by a technical support person, I then checked the timestamps and labels to make sure they were marked as finely as .25 seconds and correctly labelled. I later checked over all recordings a third time to ensure reliability in timestamping between those time-stamped early in the process and those time-stamped later. Three checks over the timestamping was probably enough to remove most error, but a fourth check by a third party may have improved the reliability of the timestamping.

Study Two analysis of the use of complete multi-word expressions suggested that the level of familiarity with the expressions prior to the intervention influenced the fluency outcomes. This observation and logical inference were based on only two instances, so the generalisability is limited. It would be informative to analyse the use of complete expressions in Study Three to investigate to what extent the expressions used

in the post-test were known before the intervention. Such analysis was outside the scope of this project; however, such an investigation would be useful to verify or disprove the hypothesis suggested from Study Two. In this section, I have described the limitations of this research and how the limitations affect the interpretation and relevance of the results. In the next section, I will suggest avenues for further research that have emerged from this project.

8.4 Future research

Future research into fluency and multi-word expressions emerging from this project can be grouped into three categories: 1) extended replication opportunities, 2) using alternative measurement methods, and 3) possible post-hoc analysis to investigate different research questions.

8.4.1 Extended replication opportunities

The most interesting future research possibility would be to replicate the study with a longer intervention period. The results of the present study left me wondering whether fluency would have been greatly improved for the experimental group over the control group had the intervention just been longer. Fluency takes time to develop, therefore, I recommend extended replication of this research whereby a longer intervention period in an EFL context is used. For example, would one semester (approximately 22 hours over 15 weeks) of intervention result in an increase in fluency over a control group?

It would also be informative to test the results of this intervention through replication in other EFL contexts. Wood (2009), Onoda (2014) and the current study have all used Japanese L1 speakers as participants. Would the same trends in results be detectable in South American, Eastern European, Asian and other EFL contexts? Also, would the various classroom activities be considered appropriate in different cultural contexts? Such investigation would help teachers in other EFL contexts to select and plan more effective classroom learning sequences and understand learner responses better.

8.4.2 Using alternative measurement methods

Procedural knowledge of multi-word expressions could potentially be directly tested using alternative methods. One potential way to measure procedural knowledge of multi-word expressions would be to do a read-a-loud test as in N. C. Ellis et al. (2008). A multi-word expression could be shown on a screen and the participant asked to read it out loud as fast as possible. Greater procedural knowledge would be indicated by shorter voice onset time from seeing the expression to uttering it and a faster articulation rate. A second potential measure also used in N. C. Ellis et al. (2008) would be to show the initial words of the expression on a screen and then show the final word which the participant would read aloud. Quicker voice onset time for the reading of the final word would also indicate greater procedural knowledge, and demonstrate the degree to which the initial words prime the final word (N. C. Ellis et al., 2008). Such measures could be a good complement or alternative to the indirect measurement of procedural knowledge of expressions through use in the role-play.

Furthermore, the ability to use multi-word expressions outside of the classroom could be measured. Natural learner interactions in the L2 outside of the classroom before and after the intervention could be recorded and analysed. Such natural spoken data could be used to measure fluency and procedural knowledge and give a more realistic picture of the impact such an intervention can have on fluency and multi-word expression use outside of the classroom.

8.4.3 Possible post-hoc analysis

Post-hoc analysis of the current results is also an interesting future research possibility. Learning of the 30 multi-word expressions was spread out over six sessions, with 10 expressions practised over two weeks as part of a themed unit. The 10 expressions from the first unit were encountered in the first two weeks, therefore by the time they were post-tested there had been a four-week delay. Following this pattern, the expressions from the second unit were post-tested with a two-week delay and the expressions from the final unit were immediately post-tested. It is possible that the post-test results may have been stronger for those expressions encountered more recently. Comparing the results for the expressions by unit would be an interesting post-hoc

investigation. Results might show whether the delay showed any effect on ability to retrieve and use the expressions. Such an analysis could reveal more about the resilience of learning/memory for the expressions to forgetting over time. Unfortunately, it was beyond the scope of enquiry for this research project.

8.5 Reflections on the PhD journey (what have I learnt)

This research project has highlighted to me the importance of sequencing activities with repeated exposure and use of target expressions (also applicable for vocabulary and grammar forms) for language learning. The importance of having many input and output activities is clear for the foreign language teaching and learning context where opportunity to use the L2 is limited beyond the classroom. Even with the limited exposure of one class a week, multiple opportunities to become familiar and comfortable with using expressions can give learners a sense of improvement and achievement; particularly when they realise they have learnt to use an expression that they did not previously know or have the ability to use. Pre- and post-course testing can help raise learner awareness of such progress and help the teacher to see what has been successfully acquired and what has not.

The collection of feedback from learners and particularly the focus groups helped me understand the learning experience from the learners' point of view. This kind of interaction between teacher and learners can be extremely helpful for teachers who want to improve their teaching skill. Understanding the cultural behaviours that come into play in the EFL context are also very important. This research project has taught me through experience that I cannot assume to know how learners feel about activities or situations and I should do my best to create space to ask for feedback from learners regarding my teaching and classroom activities.

One element of my PhD journey which will influence my own teaching is feedback. I have received a lot of feedback from my supervisors, student learning support staff and reviewers at journals. I think personally I have learnt to appreciate this critical feedback as value creating rather than a negative assessment process. I have also come to realise the motivation that positive feedback gives. Criticism is easier to utilise

when it is accompanied by praise and affirmation. Personally, I will try to keep this in mind as I give feedback to others I interact with.

The responsibility of PhD research to create new knowledge sounds intimidating, something a younger version of myself would never have imagined that I would even attempt. However, as I have lengthened in years and experience, I have realised that there are gaps in knowledge that given the resources and motivation I could attempt to fill. New knowledge helps to fill gaps in a much wider canvas that one could not hope to fill alone. I have come to understand my research contribution as filling in a few patches here in the formulaic language field and that this field is part of the wider language learning field, which itself fits into a cognition and general learning field. New research knowledge will fill old gaps and reveal new gaps in the canvas that may not have been visible previously. This thesis has helped to fill or partially fill some knowledge gaps and revealed some new gaps to investigate. If the great challenge of learning a new language is made easier by the insights revealed through this research project, then I have achieved an important objective for pursuing and completing this research project.

9 References

- Abbuhl, R. (2012). Why, when, and how to replicate research. In A. Mackey & M. G. Susan (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 296–312). Wiley-Blackwell.
- Albino, G. (2017). Improving speaking fluency in a task-based language teaching approach: The case of EFL learners at PUNIV-Cazenga. *SAGE Open*, 7(2), 2158244017691077. <https://doi.org/10.1177/2158244017691077>
- Al-Nasser, A. S. (2015). Problems of English language acquisition in Saudi Arabia: An exploratory-cum-remedial study. *Theory and Practice in Language Studies*, 5(8), 1612–1619. <http://dx.doi.org/10.17507/tpls.0508.10>
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In P. C. Anthony (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 101–122). Oxford University Press.
- Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.
- Anderson, J. R. (2000). *Learning and memory: An integrated approach* (2nd edition). John Wiley & Sons Inc.
- Anthony, L. (2014). *AntConc (Version 3.4.3) [Computer Software]*. Tokyo, Japan: Waseda University. <http://www.laurenceanthony.net/>
- Barcroft, J., & Rott, S. (2010). Partial word form learning in the written mode in L2 German and Spanish. *Applied Linguistics*, 31(5), 623–650.
- Bennett, R. (2012). Is linguistic ability variation in paired oral language testing problematic? *ELT Journal*, 66(3), 337–346. <https://doi.org/10.1093/elt/ccr066>
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275–311. <https://doi.org/10.1075/ijcl.14.3.08bib>
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263–286. <https://doi.org/10.1016/j.esp.2006.08.003>

- Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181–190). Rodopi.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405. <https://doi.org/10.1093/applin/25.3.371>
- Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercises on verb-noun collocations. *Language Teaching Research*.
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10(3), 245–261.
- Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics*, 32, 83–110. <http://dx.doi.org.helicon.vuw.ac.nz/10.1017/S0267190512000050>
- Boers, F., & Thai, C. (2017). Repeating a monologue under increasing time pressure: A replication of Thai and Boers (2016). *The TESOLANZ Journal*, 25, 1–11.
- Boersma, P., & Weenik, D. (2017). *Praat: Doing phonetics by computer* (6.0.37) [Computer software]. <http://www.praat.org/>
- Borg, S. (2010). Language teacher research engagement. *Language Teaching*, 43(4), 391–429. <https://doi.org/10.1017/S0261444810000170>
- Bruen, J., & Kelly, N. (2017). Using a shared L1 to reduce cognitive overload and anxiety levels in the L2 classroom. *The Language Learning Journal*, 45(3), 368–381. <https://doi.org/10.1080/09571736.2014.908405>
- Burns, A. (2010). *Doing action research in English language teaching: A guide for practitioners*. Routledge.
- Bybee, J. L. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–236). Routledge.

- Bybee, J. L. (2014). Usage-based grammar and second language acquisition. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure: Vol. II* (pp. 216–236). Psychology Press.
- Bybee, J. L., & Hopper, P. J. (Eds.). (2001). *Frequency and the emergence of linguistic structure* (<https://doi.org/10.1075/tsl.45>). John Benjamins. <https://doi.org/10.1075/tsl.45>
- Bygate, M. (2018). Introduction. In M. Bygate (Ed.), *Learning language through task repetition* (pp. 1–26). John Benjamins Publishing Company.
- Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language*. John Benjamins Publishing Company. <http://ebookcentral.proquest.com/lib/vuw/detail.action?docID=622981>
- Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*, 5, 31–64.
- Byrne, D. (1986). *Teaching Oral English* (2nd ed.). Longman.
- Chambers, F. (1997). What do we mean by fluency? *System*, 25(4), 535–544. [https://doi.org/10.1016/S0346-251X\(97\)00046-8](https://doi.org/10.1016/S0346-251X(97)00046-8)
- Chartrand, T. L., & van Baaren, R. (2009). Chapter 5 Human Mimicry. *Advances in Experimental Social Psychology*, Academic Press, 41, 219–274. [https://doi.org/10.1016/S0065-2601\(08\)00405-X](https://doi.org/10.1016/S0065-2601(08)00405-X)
- Cobb, T. (2020). *Compleat Web VP* (2.1) [Computer software]. <https://www.lexutor.ca/vp/comp/>
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45–61. <http://dx.doi.org.helicon.vuw.ac.nz/10.1017/S0267190512000074>
- Cook, M. (2012). Revisiting Japanese English Teachers' (JTEs) perceptions of communicative, audio-lingual, and grammar translation (yakudoku) activities: Beliefs, practices, and rationales -. *Asian EFL Journal*, 14(2), 79–98.
- Cook, V. J. (1992). Evidence for multicompetence. *Language Learning*, 42(4), 557–591. <https://doi.org/10.1111/j.1467-1770.1992.tb01044.x>
- Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations

- in speech affect language comprehension. *Cognition*, 105(3), 658–668.
<https://doi.org/10.1016/j.cognition.2006.10.010>
- Coxhead, A. (2008). Phraseology and English for academic purposes: Challenges and opportunities. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 149–161). John Benjamins Publishing.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.
[https://doi.org/10.1016/S0022-5371\(72\)80001-X](https://doi.org/10.1016/S0022-5371(72)80001-X)
- Crossley, S., & Salsbury, T. L. (2011). The development of lexical bundle accuracy and production in English second language speakers. *IRAL - International Review of Applied Linguistics in Language Teaching*, 49(1), 1–26. <https://doi.org/10.1515/iral.2011.001>
- Dang, T. N. Y., & Webb, S. (2017). Evaluating lists of high-frequency words. *ITL - International Journal of Applied Linguistics*, 167(2), 132–158. <https://doi.org/10.1075/itl.167.2.02dan>
- Davies, M. (2013). *Corpus of Contemporary American English (COCA)*. <https://corpus.byu.edu/coca/>
- Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing*. <http://journals.sagepub.com/doi/10.1177/0265532209104667>
- de Jong, N. H., Steinel, M., Florjin, A., Schoonen, R., & Hulstijn, J. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, 34(1), 5–34.
<https://doi.org/10.1017/S0272263111000489>
- de Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61(2), 533–568.
<https://doi.org/10.1111/j.1467-9922.2010.00620.x>
- Dechert, H. W. (1983). How a story is done in a second language. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 175–195). Longman.
- DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19(2), 195–221. JSTOR.
- DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second*

- language instruction* (pp. 125–151). Cambridge University Press.
- DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 94–112). Taylor & Francis.
- Ding, Y. (2007). Text memorization and imitation: The practices of successful Chinese learners of English. *System*, 35(2), 271–280. <https://doi.org/10.1016/j.system.2006.12.005>
- Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford University Press.
- Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. Dover Publications.
- Edwards, E., & Burns, A. (2016). Language teacher action research: Achieving sustainability. *ELT Journal*, 70(1), 6–15. <https://doi.org/10.1093/elt/ccv060>
- Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33–68). Cambridge University Press.
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the “phrasal teddy bear.” *Annual Review of Applied Linguistics*, 32, 17–44.
- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375–396.
- Ellis, N. C., & Wulff, S. (2015). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (pp. 75–93). Routledge.
- Ellis, R. (2005). Planning and task-based performance. In R. Ellis (Ed.), *Planning and task performance in a second language* (Vol. 11, pp. 3–36). John Benjamins Publishing.
- Ellis, R. (2015). *Understanding second language acquisition* (2nd ed.). Oxford University Press.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29–62.
- Eyckmans, J. (2009). Towards an assessment of learners’ receptive and productive syntagmatic knowledge. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language*. Palgrave Macmillan.
- Field, A. (2009). *Discovering Statistics Using SPSS*. SAGE.

- Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, 26(2), 275–301.
<https://doi.org/10.1017/S0272263104262064>
- Gasiorek, J. (2016). Theoretical perspectives on interpersonal adjustments in language and communication. In *Communication accommodation theory: Negotiating personal relationships and social identities across contexts* (pp. 13–35). Cambridge University Press.
<https://doi.org/10.1017/CBO9781316226537.002>
- Gatbonton, E., & Segalowitz, N. (1988). Creative automatization: Principles for promoting fluency within a communicative framework. *TESOL Quarterly*, 22(3), 473–492.
<https://doi.org/10.2307/3587290>
- Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes*, 61(3), 325–353.
- Giles, H. (2016). *Communication accommodation theory: Negotiating personal relationships and social identities across contexts*. Cambridge University Press.
- Glasgow, G. P., & Paller, D. L. (2014). MEXT-approved EFL textbooks and the new Course of Study. *JALT2013 Conference Proceedings*. https://jalt-publications.org/files/pdf-article/jalt2013_012.pdf
- Hagenauer, G., & Hascher, T. (2014). Early adolescents' enjoyment experienced in learning situations at school and its relation to student achievement. *Journal of Education and Training Studies*, 2(2), 20–30.
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237–258. <https://doi.org/10.1111/j.1473-4192.1994.tb00065.x>
- Herder, S., & Sholdt, G. (2014). Employing a fluency-based approach to teach the TOEFL iBT: An action research project. In T. Muller, J. Adamson, P. S. Brown, & S. Herder (Eds.), *Exploring*

- EFL fluency in Asia* (pp. 26–41). Palgrave Macmillan.
- Hirsh, D., & Coxhead, A. (2009). Ten ways of focussing on science-specific vocabulary in EAP classrooms. *English Australia Journal*, 25(1), 5–16.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Sage Publications.
- Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, 2(1). <https://doi.org/10.9707/2307-0919.1014>
- Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd ed.). McGraw Hill Professional.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207–223.
- Hosoda, Y. (2000). Teacher codeswitching in the EFL classroom. *JALT Journal*, 22.1, 69–93.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. <https://doi.org/10.1016/j.esp.2007.06.001>
- In'nami, Y., Koizumi, R., & Tomita, Y. (2020). Meta-analysis in applied linguistics. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics*. Routledge.
- Jiang, J. (2009). Designing pedagogic materials to improve awareness and productive use of L2 collocations. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language: Multiple interpretations* (pp. 99–113). Palgrave Macmillan UK.
- Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, 19(3), 357–377. <https://doi.org/10.1093/applin/19.3.357>
- Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons Inc.
- Kasahara, K. (2011). The effect of known-and-unknown word combinations on intentional vocabulary learning. *System*, 39(4), 491–499. <https://doi.org/10.1016/j.system.2011.10.001>

- Kormos, J., & Dénes, Mariann. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32(2), 145–164.
<https://doi.org/10.1016/j.system.2004.01.001>
- Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*. Pergamon Press Inc. http://www.sdkrashen.com/content/books/sl_acquisition_and_learning.pdf
- Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (Vol. 9, pp. 37–54). John Benjamins Publishing Company.
- Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). Taylor and Francis.
- Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29(4), 694–716.
<https://doi.org/10.1093/applin/amn018>
- Lavrakas, P. (2008). *Encyclopedia of survey research methods*. SAGE Publications Ltd.
- Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a likert scale. *Research in Nursing & Health*, 25(4), 295–306.
<https://doi.org/10.1002/nur.10041>
- Lennes, M. (2002). *Speech Corpus Toolkit (SpeCT)*. http://www.helsinki.fi/~lennes/praat-scripts/public/total_duration_of_labeled_segments.praat
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 387–417.
- Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggensbach (Ed.), *Perspectives on fluency* (pp. 25–42). The University of Michigan Press.
- Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Language Teaching Publications.
- Lindstromberg, Eyckmans, J., & Connebeer, R. (2016). A modified dictogloss for helping university students notice and remember conventional phrasal expressions of academic English (as L2)

- and use them in later writing. *Journal of English for Special Purposes*, 41, 12–21.
- Loewen, S., & Philp, J. (2012). Instructed second language acquisition. In A. Mackey & Susan M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 53–73). Wiley-Blackwell.
- Loewen, S., & Plonsky, L. (2016). *An A–Z of applied linguistics research methods*. Palgrave Macmillan.
- Macalister, J. (2014). Developing speaking fluency with the 4/3/2 technique: An exploratory study. *TESOLANZ Journal*, 22, 28–42.
- Macaro, E. (2005). Codeswitching in the L2 classroom: A communication and learning strategy. In E. Llurda (Ed.), *Non-native language teachers: Perceptions, challenges and contributions to the profession* (Vol. 5, pp. 63–84). Springer.
- Mackey, A. (2012). Why (or why not), when, and how to replicate research. In G. Porte (Ed.), *Replication research in applied linguistics* (pp. 21–46). Cambridge University Press.
- Mandouit, L. (2018). Using student feedback to improve teaching. *Educational Action Research*, 26(5), 755–769. <https://doi.org/10.1080/09650792.2018.1426470>
- Mártinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(3), 299–320. <https://doi.org/10.1093/applin/ams010>
- Maurice, K. (1983). The fluency workshop. *TESOL Newsletter*, 17(4), 29.
- McGuire, M., & Larson-Hall, J. (2018). Teaching formulaic sequences in the classroom: Effects on spoken fluency. *TESL Canada Journal*, 34(3), 1–25.
- McLaughlin, B. (1990). Restructuring. *Applied Linguistics*, 11(2), 113–128.
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19(6), 741–760. <https://doi.org/10.1177/1362168814567889>
- Miller G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Millett, S. (2008, April). A daily fluency programme: The key to using what you know. *Modern*

- English Teacher*, 17(2), 21-. Gale Academic OneFile.
- Morita, L. (2015). English, language shift and values shift in Japan and Singapore. *Globalisation, Societies and Education*, 13(4), 508–527. <https://doi.org/10.1080/14767724.2014.967184>
- Myles, F. (2004). From data to theory: The over-representation of linguistic knowledge in SLA. *Transactions of the Philological Society*, 102(2), 139–168. <https://doi.org/10.1111/j.0079-1636.2004.00133.x>
- Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, 48(3), 323–364. <https://doi.org/10.1111/0023-8333.00045>
- Nakata, T. (2013). *Optimising second language vocabulary learning from flashcards* [Unpublished doctoral dissertation, Victoria University of Wellington]. <http://restrictedarchive.vuw.ac.nz/handle/123456789/8568>
- Nakata, T. (2016). Effects of retrieval formats on second language vocabulary learning. *International Review of Applied Linguistics in Language Teaching*, 54(3), 257–289. <https://doi.org/10.1515/iral-2015-0022>
- Nation, I. S. P. (1989). Improving speaking fluency. *System*, 17(3), 377–384. [https://doi.org/10.1016/0346-251X\(89\)90010-9](https://doi.org/10.1016/0346-251X(89)90010-9)
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press.
- Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning & Teaching*, 1(1), 2–13. <https://doi.org/10.2167/illt039.0>
- Nation, I. S. P. (2012). *The BNC/COCA headword lists*. <https://www.victoria.ac.nz/lals/about/staff/paul-nation#vocab-lists>
- Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd Edition). Cambridge University Press.
- Nation, I. S. P. (2014). Developing Fluency. In T. Muller, J. Adamson, P. S. Brown, & S. Herder (Eds.), *Exploring EFL fluency in Asia* (pp. 11–25). Palgrave Macmillan.
- Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. Routledge.

- Nation, I. S. P., & Yamamoto, A. (2012). Applying the four strands. *International Journal of Innovation in English Language Teaching and Research*, 1(2), 167–181.
- Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.
- Nishino, T., & Watanabe, M. (2008). Communication-oriented policies versus classroom realities in Japan. *TESOL Quarterly*, 42(1), 133–138. <https://doi.org/10.2307/40264432>
- Nunan, D., & Richards, J. C. (Eds.). (2015). *Language Learning Beyond the Classroom*. Taylor and Francis. <http://www.myilibrary.com?id=719146>
- Onoda, S. (2014). An exploration of effective teaching approaches for enhancing the oral fluency of EFL students. In T. Muller, J. Adamson, P. S. Brown, & S. Herder (Eds.), *Exploring EFL fluency in Asia* (pp. 120–142). Palgrave Macmillan.
- Pallant, J. (2007). *SPSS survival manual: A step-by-step guide to data analysis using SPSS for Windows*. Open University Press.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–225). Longman.
- Peters, A. M. (1983). *The units of language acquisition*. Cambridge University Press.
- Porte, G. (2012). *Replication research in applied linguistics*. Cambridge University Press.
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- Rogers, J., & Révész, A. (2020). Experimental and quasi-experimental designs. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics*. Routledge.
- Salkind, N. (2010). *Encyclopedia of research design*. SAGE Publications Ltd. <http://methods.sagepub.com/reference/encyc-of-research-design/n44.xml>
- Samaranayake, S. W. (2016). Oral competency of ESL/EFL learners in Sri Lankan rural school context. *SAGE Open*, 6(2), 2158244016654202. <https://doi.org/10.1177/2158244016654202>
- Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, 6(2), 72–

78. [https://doi.org/10.1016/0346-251X\(78\)90027-1](https://doi.org/10.1016/0346-251X(78)90027-1)
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–33). Cambridge University Press.
- Schmitt, N., Dornyei, Z., Apolphs, S., & Durow, V. (2004). Knowledge and acquisition of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences acquisition, processing, and use* (Vol. 9, pp. 55–86). J. Benjamins. <http://public.eblib.com/choice/publicfullrecord.aspx?p=622575>
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing*, 18(1), 55–88. <https://doi.org/10.1191/026553201668475857>
- Scrivener, J. (2005). *Learning teaching: The essential guide to English language teaching* (Second). Macmillan Education.
- Shadish, W. R., & Clark, M. H. (2004). Comparison group. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), *The SAGE encyclopedia of social science research methods* (Vol. 1, pp. 154–155). SAGE Publications, Inc.
- Shiki, O., Mori, Y., Kadota, S., & Yoshida, S. (2010). Exploring differences between shadowing and repeating practices: An analysis of reproduction rate and types of reproduced words. *ARELE: Annual Review of English Language Education in Japan*, 21, 81–90.
- Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal: English Language Teaching Journal*, 62(4), 339–348. <https://doi.org/10.1093/elt/ccm091>
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487–512. <https://doi.org/10.1093/applin/amp058>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology*, 37(3), 776–784.
- Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*,

- 36(5), 549–569. <https://doi.org/10.1093/applin/amt054>
- Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (2019). Formulaic language: Setting the scene. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 1–15). Routledge.
- Siyanova-Chanturia, A., & Van Lanker Sittis, D. (2019). Online processing of formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 38–61). Routledge.
- Skehan, P. (1996a). Second-language acquisition research and task-based instruction. In J. Willis & D. Willis (Eds.), *Challenge and change in language teaching* (pp. 17–30). Heinemann.
- Skehan, P. (1996b). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17(1), 38–62. <https://doi.org/10.1093/applin/17.1.38>
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510–532. <https://doi.org/10.1093/applin/amp047>
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93–120. <https://doi.org/10.1111/1467-9922.00071>
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 183–205). Cambridge University Press.
- Snellings, P., Gelderen, A. van, & Glopper, K. de. (2002). Lexical retrieval: An aspect of fluent second-language production that can be enhanced. *Language Learning*, 52(4), 723–754. <https://doi.org/10.1111/1467-9922.00202>
- Snoder, P., & Reynolds, B. L. (2019). How dictogloss can facilitate collocation learning in ELT. *ELT Journal*, 73(1), 41–50. <https://doi.org/10.1093/elt/ccy024>
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, 36(2), 139–152. <https://doi.org/10.1080/09571730802389975>
- Stangroom, J. (2018). *Effect Size Calculator for T-Test*. <https://www.socscistatistics.com/effectsize/>
- Stengers, H., Boers, F., Housen, A., & Eyckmans, J. (2010). Does “chunking” foster chunk-uptake? In S. De Knop, F. Boers, & De Rycker, A. (Eds.), *Fostering language teaching efficiency through*

cognitive linguistics (pp. 99–117). De Gruyter Mouton.

- Storch, N., & Wigglesworth, G. (2003). Is there a role for the use of the L1 in an L2 setting? *TESOL Quarterly*, 37(4), 760–770. <https://doi.org/10.2307/3588224>
- Taguchi, N. (2007). Chunk learning and the development of spoken discourse in a Japanese as a foreign language classroom. *Language Teaching Research*, 11(4), 433–457. <https://doi.org/10.1177/1362168807080962>
- Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *IRAL: International Review of Applied Linguistics in Language Teaching*, 54(2), 133–150. <https://doi.org/10.1515/iral-2016-9994>
- Tavakoli, P., Campbell, C., & McCormack, J. (2016). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly*, 50(2), 447–471. <https://doi.org/10.1002/tesq.244>
- Tavakoli, P., & Hunter, A.-M. (2018). Is fluency being ‘neglected’ in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, 22(3), 330–349. <https://doi.org/10.1177/1362168817708462>
- Tavakoli, P., & Uchihara, T. (2020). To what extent are multiword sequences associated with oral fluency? *Language Learning*. <https://doi.org/10.1111/lang.12384>
- Thai, C., & Boers, F. (2016). Repeating a monologue under increasing time pressure: Effects on fluency, complexity, and accuracy. *TESOL Quarterly*, 50(2), 369–393. <https://doi.org/10.1002/tesq.232>
- Thomson, H. (2018). Building speaking fluency with multiword expressions. *TESL Canada Journal*, 34(3), 26–53. <https://doi.org/10.18806/tesl.v34i3.1272>
- Thomson, H., Boers, F., & Coxhead, A. (2019). Replication research in pedagogical approaches to spoken fluency and formulaic sequences: A call for replication of Wood (2009) and Boers, Eyckmans, Kappel, Stengers & Demecheleer (2006). *Language Teaching*, 52(3), 406–414. <https://doi.org/10.1017/S0261444817000374>
- Tian, L., & Macaro, E. (2012). Comparing the effect of teacher codeswitching with English-only

- explanations on the vocabulary acquisition of Chinese university students: A lexical focus-on-form study. *Language Teaching Research*, 16(3), 367–391.
<http://dx.doi.org/helicon.vuw.ac.nz/10.1177/1362168812436909>
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17(1), 84–119. <https://doi.org/10.1093/applin/17.1.84>
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569–613. <https://doi.org/10.1111/j.1467-9922.2010.00622.x>
- Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, 69(3), 559–599. <https://doi.org/10.1111/lang.12343>
- Ullman, M. T. (2015). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 135–158). Taylor & Francis.
- Ullman, M. T., & Lovelett, J. T. (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research*, 34(1), 39–65. <https://doi.org/10.1177/0267658316675195>
- Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye movement study into the processing of formulaic language. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (Vol. 9, pp. 153–168). John Benjamins Publishing Company.
- VanPatten, B., & Williams, J. (2015). Early theories in SLA. In B. VanPatten & J. Williams (Eds.), *Theories in Second Language Acquisition: An Introduction* (pp. 17–33). Routledge.
- Villavicencio, F. T., & Bernardo, A. B. I. (2013). Positive academic emotions moderate the relationship between self-regulation and academic achievement. *The British Journal of Educational Psychology*, 83(Pt 2), 329–340. <https://doi.org/10.1111/j.2044-8279.2012.02064.x>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

- Wajnryb, R. (1990). *Grammar dictation*. Oxford University Press.
- Walker, I., & Utsumi, T. (2006). Memorizing dialogues: The case for performance exercises. In W. M. Chan, K. N. Chin, & T. Suthiwan (Eds.), *Foreign language teaching in Asia and beyond: Current perspectives and future directions* (pp. 243–269). Centre for Language Studies.
- Wang, R., Hempton, B., Dugan, J. P., & Komives, S. R. (2008). Cultural differences: Why do Asians avoid extreme responses? *Survey Practice*, 1(3).
<http://www.surveypractice.org/index.php/SurveyPractice/article/view/224>
- Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46–65. <https://doi.org/10.1093/applin/aml048>
- Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL Quarterly*, 43(1), 55–77. JSTOR.
- Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91–120. <https://doi.org/10.1111/j.1467-9922.2012.00729.x>
- Wennerstorm, A. (2000). The role of intonation in second language fluency. In H. Riggensbach (Ed.), *Perspectives on fluency* (pp. 102–127). The University of Michigan Press.
- Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics (Oxford)*, 32(4), 430–449.
- Wood, D. (2001). In search of fluency: What is it and how can we teach it? *Canadian Modern Language Review*, 57(4), 573–589. <https://doi.org/10.3138/cmlr.57.4.573>
- Wood, D. (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *The Canadian Journal of Applied Linguistics*, 12(1), 39–57.
- Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. Continuum.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Wray, A. (2004). “Here’s one I prepared earlier”: Formulaic language learning on television. In N. Schmitt (Ed.), *Formulaic sequences acquisition, processing, and use* (pp. 249–268). J.

Benjamins.

- Wray, A. (2019). Concluding question: Why don't second language learners more proactively target formulaic sequences? In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 248–269). Routledge.
- Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorised language as a gauge of nativelike competence. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 123–147). John Benjamins Publishing Company.
- Wulff, S. (2019). Acquisition of formulaic language from a usage-based perspective. In A. Siyanova-Chanturia & A. Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 19–37). Routledge.
- Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, 44(4), 647–668.
- Yu, X. (2009). A formal criterion for identifying lexical phrases: Implication from a classroom experiment. *System*, 37(4), 689–699.
- Zhang, X. (2017). Effects of receptive-productive integration tasks and prior knowledge of component words on L2 collocation development. *System*, 66, 156–167.
<https://doi.org/10.1016/j.system.2017.03.019>
- Zhang, Y. (2009). Discussing the extended writing project. *ELT Journal*, 63(2), 167–169.
<https://doi.org/10.1093/elt/ccn071>

Appendix 1 Ethics approval from Japanese institution

(様式3)

承認番号	H27-07-S07
------	------------

審 査 結 果 通 知 書

平成 2 7 年 1 2 月 2 2 日

研究責任者 殿

課題名：Developing fluency using multi-word expressions

上記課題の研究実施計画を平成 2 7 年 1 2 月 1 7 日の委員会で審査し、下記のとおり判定しましたので、通知します。

記

判 定	承認
条件又は変更勧告 の内容及び理由等	※委員会から別添の意見が付されていますので、参照ください。

Appendix 2 Ethics approval from New Zealand institution



Phone [REDACTED]
Email [REDACTED]

MEMORANDUM

TO	Haidee Thomson
COPY TO	Averil Coxhead
FROM	[REDACTED], Convener, Human Ethics Committee
DATE	24 November 2015
PAGES	1
SUBJECT	Ethics Approval: 22354 Developing fluency with multi-word expressions

Thank you for your application for ethical approval, which has now been considered by the Standing Committee of the Human Ethics Committee.

Your application has been approved from the above date and this approval continues until 30 September 2018. If your data collection is not completed by this date you should apply to the Human Ethics Committee for an extension to this approval.

Best wishes with the research.

Kind regards

[REDACTED]

[REDACTED]
Convener, Victoria University Human Ethics Committee

[REDACTED] Personal details redacted for privacy reasons

Appendix 3 First amendment to ethics approval



Phone [REDACTED]
Email [REDACTED]

MEMORANDUM

TO	Haidee Thomson
COPY TO	Averil Coxhead
FROM	[REDACTED], Convener, Human Ethics Committee
DATE	7 January 2016
PAGES	1
SUBJECT	Ethics Approval: 22354 Developing fluency with multi-word expressions

Thank you for your request to amend your ethics approval. This has now been considered and the request granted.

Your application has approval until 30 September 2018. If your data collection is not completed by this date you should apply to the Human Ethics Committee for an extension to this approval.

Best wishes with the research.

Kind regards

[REDACTED]

[REDACTED]
Convener, Victoria University Human Ethics Committee

[REDACTED] Personal details redacted for privacy reasons

Appendix 4 Second amendment to ethics approval



Phone [redacted]
Email [redacted]

MEMORANDUM

TO	Haidee Thomson
FROM	[redacted], Convenor, Human Ethics Committee
DATE	14 October 2017
PAGES	1
SUBJECT	Ethics Approval Number: 22354 Title: Developing fluency with multi-word expressions

Thank you for your application to amend your ethics approval. This has now been considered and the request granted.

In the case of an amendment, this approval is valid until the end date of your original ethics approval; in the case of an extension, this approval applies until the new end date that you have nominated. If your data collection is not completed by this date you should apply to the Human Ethics Committee for an extension to this approval.

Best wishes with the research.

Kind regards,

[redacted signature]

[redacted]
Convenor, Victoria University of Wellington Human Ethics Committee

[redacted] Personal details redacted for privacy reasons

Appendix 5 Learner information and consent page (English translation)



English Language Teaching Methods Research

I am conducting research into the effectiveness of English language teaching methods as part of our class activities, please consider this request for participation in this research.

-PARTICIPANT INFORMATION SHEET-

- Participation requires no extra effort or time from you outside of class.
- Your level of participation will in no way influence the relationship with the teacher, or your course assessment.
- Test results and survey feedback related to the class activities in this class will be added to a data set which will be statistically analysed for overall trends.
- The information from this research will only be used by the research team, in order to help deepen understanding about language education.
- Results will be shared in conference presentations and academic reports.
- Participants will not be named or identified in any reports.
- All research data will be kept securely and destroyed 5 years after the research ends.

There is no obligation to consent, but, if you do, you can:

- withdraw from the study up until December 15, 2015
- ask any questions about the study at any time
- read any reports of this research by emailing me to request a copy

After reading this information sheet you can decide whether or not to take part in this research. If you decide to participate, thank you. If you decide not to take part, thank you for considering my request.

If you have any questions, either now or in the future, please feel free to contact:

Researcher: Haidee Thomson (Japanese OK)

Institute: Muroran Institute of Technology

Contact details redacted for privacy reasons

Supervising researcher: Dr. Averil Coxhead (English only)

Institute: Victoria University of Wellington,

Institute address: Kelburn Parade, Wellington, New Zealand

Contact details redacted for privacy reasons

Human Ethics Committee information

This research is being carried out in collaboration with Victoria University of Wellington. The project has been approved both by Muroran Institute of Technology Human Ethics Committee and the Victoria University of Wellington Human Ethics Committee (No. 22354).

If you have any concerns about the ethical conduct of the research you may contact Professor *** from Muroran Institute of Technology or Associate Professor *** from Victoria University of Wellington.

- Professor [REDACTED]: (Japanese OK)
- [REDACTED] (Victoria University HEC Convener) (English only)

[REDACTED] Personal details redacted for privacy reasons

PARTICIPANT consent

I consent for my test results and survey feedback on activities in class to be analysed and used in this study.

☐ Consent

☐ Do not consent

I have read and understood the above information

Name

Student number

Appendix 6 Learner information and consent page (Japanese)



英語指導方法に関する研究

私(トムソン)は、授業の一環として、英語の指導方法の有効性について調査を行っており、皆様にこの調査にご協力いただきたいと考えております。

ー調査の事前情報ー

- この段階への参加者には、授業以外の労力と時間をかけていただくことはありません。
- 調査への協力の度合いは、教員との関係または教科の成績にいかなる影響をも与えることはありません。
- 質問紙などへの回答は集計して全体的な傾向を統計的に分析いたします。
- 調査によって知れた情報は研究チームでのみ使用され、言語の教育についての理解を深めるために利用されます。
- 結果は、学会での発表および論文や報告書という形で公表されます。
- いかなる報告書においても、参加者名が特定されることはありません。
- 調査データは、調査終了後 5 年間安全に保管した後、破棄されます。

皆様にはこの申し出を承諾する義務はありませんが、もし協力していただける場合には、以下のようなことが可能です。

- 2015 年 12 月 15 日までに調査への協力を取りやめること。
- 調査に関していつでも、いかなる疑問についても質問すること。
- この調査のいかなる報告書もお読みいただくこと(私宛の E メールにて報告書をご依頼ください)

上記の事前情報を一読いただいた上で、調査にご協力いただけるか否かをお決めください。ご協力いただける方には、心からお礼申し上げます。参加していただけない皆様にも、この依頼についてご検討いただきましたことに感謝いたします。

現在または今後、ご質問がありましたら、お気軽に下記の者までお問い合わせください。

研究者:トムソン・ヘイディ(日本語可)

所属:室蘭工業大学

Contact details redacted for privacy reasons

指導研究者:Dr. Averil Coxhead(英語のみ)

所属:Victoria University of Wellington

所属住所:Kelburn Parade, Wellington, New Zealand


Contact details redacted for privacy reasons

ヒト倫理委員会に関するお知らせ

- この調査は、ヴィクトリア大学ウェリントンとの協力の下で行われます。このプロジェクトは、室蘭工業大学およびヴィクトリア大学ウェリントンの承認(番号 0000022354)を受けております。

当研究について倫理上の懸念をお持ちの方は、室蘭工業大学の***教授又はヴィクトリア大学ウェリントンの***にご連絡ください。

-  教授
-  (Victoria University HEC Convener) (英語のみ)

 Personal details redacted for privacy

私は、私のテスト結果と授業中のアクティビティに関する質問票への回答が、分析され本研究に利用されることに同意します。

- ☒ 同意する
- ☐ 同意しない

私は上記の事前情報を読んでおり、理解しています。

氏名

学生番号

Appendix 7 Participant information sheet and consent form for the focus group (English translation)



English Language Teaching Methods Research

Focus group PARTICIPANT INFORMATION SHEET

Thank you for your contribution to this research project into the effectiveness of English language teaching methods. I would like to invite you participate further in a focus group where participants will talk with others in the group about impressions of the class activities. This discussion will be audio-recorded. Participation in this way will require about 30 minutes of your time outside of class and the discussion will be in Japanese.

This research is confidential. Participants will not be identified in any reports. All research data (focus group audio-recordings and transcripts) will be kept securely and destroyed 5 years after the research ends. Your level of participation and anything you say as part of the focus group will in no way impact the relationship with the teacher, or on learning and assessment.

The information from this research will be used to help other teachers and researchers to understand more about language teaching and learning. Results will be shared in conference presentations and academic reports.

You do not have to accept this invitation. But, if you do decide to participate, you can:

- ask any questions about the study at any time
- withdraw from this study up until March 31, 2016 (focus group recordings cannot be withdrawn due to impracticalities).
- read any reports of this research by emailing me to request a copy
- receive 797 yen into your bank account

Human Ethics Committee information

This research is being carried out in collaboration with Victoria University of Wellington. The project has been approved both by Muroran Institute of Technology

Human Ethics Committee and the Victoria University of Wellington Human Ethics Committee (No. 22354).

If you have any questions, either now or in the future, please feel free to contact:

Researcher:

Name: Haidee Thomson

Contact details redacted for privacy reasons

Supervising researcher:

Name: Dr. Averil Coxhead

Role: Senior Lecturer

School: Victoria University of Wellington,
Kelburn Parade, Wellington, New Zealand

Contact details redacted for privacy reasons

If you have any concerns about the ethical conduct of the research you may contact

Muroran Institute of Technology: Professor [REDACTED]

[REDACTED] Personal details redacted for privacy

or

Victoria University HEC Convener: Associate Professor [REDACTED]

[REDACTED] Personal details redacted for privacy



English Language Teaching Methods Research

Focus group PARTICIPANT CONSENT FORM

I would like to take part in a focus group to discuss the classroom activities (this will require about 30 minutes of my time outside of class, discussion will be in Japanese). I agree to have this group discussion audio-recorded and included in the study. ☐ Yes ☐ No

I understand that:

- My level of participation will in no way impact the relationship with the teacher, or on learning and assessment
- Nothing I say as part of the focus group will influence the relationship with the teacher, or on learning and assessment.
- I may withdraw from this study up until March 31, 2016 (focus group recordings cannot be withdrawn due to impracticalities).
- I can ask questions about the research at any time.
- Any information I provide will be kept confidential to the researchers.
- My name or identifying information will not be used in any reports or presentations.
- I understand that the summary of results will be presented at conferences, and used in academic reports.
- I can receive 797 yen to my bank account

Name of participant: _____

Date: _____

Appendix 8 Participant information sheet and consent form for the focus group (Japanese)



英語指導方法に関する調査

フォーカス・グループ協力者の皆様への事前情報

今まで指導方法の有効性について調査をご協力いただき、心から感謝します。皆様にこの調査にご協力いただきたいと考えております。

さらに、クラスでのアクティビティの印象についてグループで話し合うフォーカス・グループに参加することを選択していただくことができます。この話し合いは、録音させていただきます。こちらに参加してくださる方には、授業以外に30分ほどお時間をいただくことになり、話し合いは日本語で行われます。

この調査で得られた情報は、研究チームでのみ使用されます。またいかなる報告書においても、参加者名が特定されることはありません。調査データの全て（フォーカス・グループでの録音音声やその書き起こし文書）は、調査終了後5年間安全に保管した後、破棄されます。調査への協力の度合いは、教員との関係または教科の成績にいかなる影響をも与えることはありません。

この調査によって知りえた情報は、教師や研究者が言語の教育や学習についての理解を深めるために利用されます。結果は、学会での発表および論文や報告書という形で公表されます。

皆様にはこの申し出を承諾する義務はありませんが、もし協力していただける場合には、以下のようなことが可能です。

- 調査に関していつでも、いかなる疑問についても質問すること
- 2016年03月31日までに調査への協力を取りやめること（ただし、フォーカス・グループのデータは手続き上残ります）

- この調査のいかなる報告書もお読みいただくこと（私宛のEメールにてコピーの送付をご依頼ください）
- 銀行口座に797円が支払われます。

ヒト倫理委員会に関するお知らせ

この調査は、ヴィクトリア大学ウェリントンとの協力の下で行われます。
このプロジェクトは、室蘭工業大学およびヴィクトリア大学ウェリントンの承認（番号 22354）を受けております。

現在または今後、ご質問がありましたら、お気軽に下記の者までお問い合わせください。

研究者: トムソンヘイディ
室蘭工業大学

Contact details redacted for privacy reasons

指導研究者 Dr. Averil Coxhead
Victoria University of Wellington,
Kelburn Parade, Wellington,
New Zealand

Contact details redacted for privacy reasons

当研究について倫理上の懸念をお持ちの方は、室蘭工業大学の***教授にご連絡ください。メール***又は

Victoria University HEC Convener: Associate Professor ■■■

■■■ Personal details redacted for privacy

English Language Teaching Methods Research

グループディスカッション協力者の同意書

私は、授業中のアクティビティに関する話し合いに参加
したいと思います（話し合いは 30 分程度をかけて、講義時間
以外の時間を使い、日本語で行います）。その場でのグループ
ディスカッションが録音され、研究に利用されることにも同意
します。

Yes ☐ No ☐

私は以下のことを理解しています。

- 調査への協力の度合いは、教員との関係または教科の成績に全く影響しないこと
- フォーカス・グループの発言に関しては、教員との関係または教科の成績に全く影響しないこと
- どの時点においても、この調査に関するいかなる疑問でも、質問できること
- 私から得た情報は、研究チームにのみ使用されること
- 2016 年 03 月 31 日までに調査への協力を取りやめること（ただし、フォーカス・グループのデータは手続き上残ります）
- 報告書または発表においても、参加者名が特定されることはありません
- 調査結果の要約は学会で発表され、論文や報告書に利用されること
- 銀行口座に 7 9 7 円が支払われます。

参加者の署名と学生番号:

日付: _____

Appendix 9 Model dialogues

Model dialogues and vocabulary profile (Study One, Café)

P: Right, Jon. *We have about an hour* for lunch. The next meeting is at two thirty.

J: That's good. *I usually only have* time for a quick lunch, often at my desk!

P: Let's see what they *have on the menu*. *What would you like* Jon?

J: Mmm the salmon looks good, *but so does the omelette*. *What are you going to get*?

P: *I think I'll* have the chicken.

J: OK, *I'll have the* salmon.

P: Shall we get a salad to share?

J: Great idea, and *are you going to get a drink*?

P: Mm, I think I'll get a coffee, how about you?

J: Coffee, yeah me too.

P: OK, let's order!

Vocabulary profile (VP Compleat Input BNC_COCA families)

K1: 100 tokens

K2: 2 tokens (desk, salad)

K3: 1 token (menu)

K4: 2 tokens (salmon)

Proper nouns, cognates (jon, omelette)

verbal fillers not counted (mm, mmm)

recategorized as K1 (ok)

Model dialogue and vocabulary profile (Study One, Directions)

A: Excuse me, *how do I get* to the National Museum? *Can you tell me* how?

B: The National Museum, um...go straight along this road, *when you get to* the traffic lights, turn left onto Green street.

A: *Turn left at the* traffic lights onto Green street, OK.

B: Then, walk along Green street *until you get to* the Square. *You should see the* Museum on the right. The Museum is on the *right hand side of* the Square.

A: So, I turn left at the traffic lights then go up Green Street *until I get to* the Square.

B: Yes, and the Museum *is on the right* side of the Square.

A: OK. Thank you very much.

B: No worries, *have a good day!*

A: Thank you! You too!

Vocabulary profile (VP Compleat Input BNC_COCA families)

K1: 113 tokens

K2: 3 tokens (traffic)

K3: 5 tokens (museum)

verbal fillers not counted (um)

recategorized as K1 (ok)

Model dialogue and vocabulary profile (Study One, Hotel)

Staff: Welcome to the _____ Hotel. How may I help you?

C: I'd like a room please.

Staff: *How many are in your party?*

C: Just two.

Staff: What type of room would you like? We have doubles and twins available.

C: I'd like a double, please.

Staff: OK, how many nights would you like to stay?

C: Sorry, *could you repeat that* please?

Staff: Certainly. How many nights would you like to stay?

C: Oh, just tonight.

Staff: OK. May I have your name, please?

C: Timothy Findley.

Staff: *How do you spell* that please?

C: F-I-N-D-E-...

Staff: Sorry, can you repeat that please?

C: F-I-N-D-L-E-Y

Staff: OK, um that will be \$120. *How will you be* paying?

C: Is Visa OK?

Staff: *That will be fine.*

C: Is breakfast included?

Staff: No, *it is an extra* \$15 each. *Would you like to* add it your reservation?

C: Um, *let me think about* it. Does the room have WIFI?

Staff: Yes, *the details are in* your room in the folder on the desk.

C: OK, thank you.

Staff: Do you have any other questions?

C: Not at this stage, thank you.

Staff: OK, you're in room 405 on the fourth floor. Here's your key. Enjoy your stay.

C: *Thank you very much.*

Vocabulary profile (VP Complete Input BNC_COCA families)

K1: 174 tokens

K2: 12 tokens (welcome, twins, available, repeat, spell, dollars, included, reservation, details, desk)

K5: 1 token (visa)

K6: 1 token (folder)

Proper nouns, cognates (Timothy Findley, WIFI)

verbal fillers not counted (um, oh)

recategorized as K1 (ok)

Model dialogue and vocabulary profile (Study Two & Three, Hotel)

Staff: Good afternoon. How may I help you?

C: I'd like a room please.

Staff: How many are in your party?

C: Just two.

Staff: What type of room would you like?

C: I'd like a double, please.

Staff: OK, how many nights would you like to stay?

C: Just tonight.

Staff: OK. May I have your name, please?

C: Timothy Findley.

Staff: How do you spell that please?

C: F-I-N-D-E-...

Staff: Sorry, could you repeat that please?

C: F-I-N-D-L-E-Y

Staff: OK, that will be \$120. How will you be paying?

C: Is Visa OK?

Staff: That will be fine.

C: Is breakfast included?

Staff: No, it is an extra \$15 each. Would you like to add it your reservation?

C: Um, let me think about it. Does the room have WIFI?

Staff: Yes, the details are in your room.

C: OK, thank you.

Staff: OK, you're in room 405 on the fourth floor. Here's your key. Enjoy your stay.

C: Thank you very much.

K1: 136 tokens 94.4%

K2: 7 tokens 99.3% (repeat, spell, dollars, included, reservation, details)

K5: 1 token 100% (visa)

Proper nouns, cognates (Timothy Findley, WIFI)

verbal fillers not counted (um, oh)

recategorized as K1 (ok)

Appendix 10 The 30 multi-word expressions with frequency and function information

		Unit	Target multi-word expression	Frequency profile of vocabulary	Occurrences within COCA spoken WORDS corpus (95,565,075)	Occurrences per million words (COCA)	Function
1	Café		we have about an	K1	59	0.62	Information
2	Café		I usually only have	K1	0 in COCA, but 1,200,000,000 hits in a Google Search		Information
3	Café		have on the menu	have on the (K1) menu (K2)	5	0.05	Information
4	Café		What would you like	K1	315	3.30	Question
5	Café		but so does the	K1	7	0.07	Comparison
6	Café		what are you going	K1	1075	11.25	Question
7	Café		I think I will	K1	68	0.71	Information
8	Café		I will have the	K1	12	0.13	Information
9	Café		are you going to	K1	4207	44.02	Question

10	Café	to get a drink	K1	18	0.19	Question
1	Directions	How do I get	K1	78	0.82	Question
2	Directions	Can you tell me	K1	481	5.03	Question
3	Directions	When you get to	K1	264	2.76	Instruction
4	Directions	turn left at the	K1	3	0.03	Instruction
5	Directions	until you get to	K1	33	0.35	Instruction
6	Directions	You should see the	K1	32	0.33	Instruction
7	Directions	right hand side of	K1	7	0.07	Information
8	Directions	until I get to	K1	10	0.10	Information
9	Directions	is on the right	K1	67	0.70	Information
10	Directions	have a good day	K1	149	1.56	Greeting
1	Hotel	How many are in	K1	9	0.09	Question
2	Hotel	could you repeat that	could you that (K1) repeat (K2)	14	0.15	Question
3	Hotel	How do you spell	How do you (K1)	13	0.14	Question

			spell (K2)			
4	Hotel	How will you be	K1	11	0.12	Question
5	Hotel	That will be fine	K1	7	0.07	Information
6	Hotel	It is an extra	K1	3	0.03	Information
7	Hotel	Would you like to	K1	1145	11.98	Question
8	Hotel	let me think about	K1	21	0.22	Response
9	Hotel	the details are in	the, are, in (K1) details (K2)	3	0.03	Information
10	Hotel	Thank you very much.	K1	11859	124.09	Response

Appendix 11 Language learning survey (English translation)

1) Gender (this question was added for Study Two & Three)

Female

Male

2) Are you an international student? Yes, No

The following question only appeared if YES was selected to the previous question

Please advise your nationality. Please advise how many years you have studied English.

2.1) Nationality:

2.2) Years studying English:

3) How many courses with Ms. Thomson do you have this semester? 1, 2

Please answer these questions about your English language learning background.

4) Have you ever attended a private English conversation school or had private English lessons? Yes, No

The following question only appeared if YES was selected to the previous question

4.1) How long did you attend? (0-6 months, 6-12 months, 12-18 months, 18-24 months, 2-3 years, 3-4 years, 4-5 years, more than 5 years)

The following question only appeared if YES was selected to the previous question

4.2) Are you currently attending a private English conversation school? Yes, No

5) Have you ever visited other countries where you heard and spoke English every day?
Yes, No

The following question only appeared if YES was selected to the previous question

5.1) How long were you in countries where you heard and spoke English every day?
(1-14 days, 15-30 days, 1-2 months, 2-3 months, more than 3 months)

6) Have you ever studied English in another country? Yes, No

The following question only appeared if YES was selected to the previous question

6.1) Which country?

7) What is your purpose for learning English? (To get course credit, To improve my speaking & listening ability, To improve my writing and reading ability, To be able to communicate with foreigners using English, To improve my TOEIC score, To improve my employability)

8) What is your major? (Department of Civil Engineering and Architecture, Department of Mechanical, Aerospace, and Materials Engineering, Department of Applied Sciences, Department of Information and Electronic Engineering)

9) What was your most recent TOEIC score? 。

10-250

255-400

405-600

605-780

785-900

905-990

Appendix 12 Language learning survey in Japanese

1) 性別 女子 / 男子

2) 留学生ですか。はい / いいえ

The following question only appeared if YES was selected to the previous question

国籍を教えてください。英語は何年間を勉強したことがありますか。

2.1) 国籍

2.2) 英語学習の期間 (年間)

3) 今学期の授業はトムソと何個受講していますか。1 / 2

あなた、自身の英語学習歴について教えてください。

4) 英会話学校に通ったり、マンツーマンの英語教育を受けたりしたことがありますか。はい / いいえ

The following question only appeared if YES was selected to the previous question

4.1) どのくらい通っていましたか。

0 - 6 ヶ月

6 - 12 ヶ月

12 - 18 ヶ月

18 - 24 ヶ月

2 - 3 年間

3 - 4 年間

4 - 5 年間

5 年以上

The following question only appeared if YES was selected to the previous question

4.2) 今でも英会話学校に通っていますか。 はい / いいえ

5) 英語を日常的に話したり聞いたりする国に行ったことがありますか。

はい / いいえ

The following question only appeared if YES was selected to the previous question

5.1) 英語を日常的に話したり聞いたりする国に行った期間はどのくらいですか。

1 - 1 4 日間

1 5 - 3 0 日間

1 - 2 ヶ月間

2 - 3 ヶ月間

3 ヶ月間以上

6) 海外で英語を勉強したことありますか。 はい / いいえ

The following question only appeared if YES was selected to the previous question

6.1) 国の名前教えてください。

7) 英語を勉強する主な目的は何ですか。

単位を取得するため

英語を話したり聞いたりする能力を上達させるため

英語を書いたり読んだりする能力を上達させるため

外国人と英語でコミュニケーションができるようにするため

TOEICスコアを上げさせるため

就職確率を高めるため

8) 専攻を教えてください。

建築社会基盤系学科 (1)

機械航空創造系学科 (2)

応用理化学系学科 (3)

情報電子工学系学科 (4)

9) 最新のTOEICスコアを教えてください。

- ☐ 10-250
- ☐ 255-400
- ☐ 405-600
- ☐ 605-780
- ☐ 785-900
- ☐ 905-990

Appendix 13 Pre- and post-intervention dialogue instructions for Study One.

Record your conversation

こちらのロールプレイの結は、教科の成績にいかなる影響をも与えることはありません。ご安心ください。友達とカフェに行って、メニューについて英語のみで話しあってください。

You are at a restaurant for lunch with your friend

- *Try to talk in English only*
- *You have 1 hour to order and eat*
- *Talk about the menu with your partner*
- *Ask what they want to order*
- *Decide together what you will order*

Here is a translation of the Japanese above:

The results from this role-play will not affect your grade. Don't worry.

Go to the café with your friend, discuss the menu in English only.

Appendix 14 Café menu (realia) that participants discussed for the dialogue

<h1>Olive Lunch Menu 11am-2pm</h1>	
Salmon \$15	Grilled salmon served with baked potatoes and butter
Chicken \$15	Roast chicken served with roast potatoes and butter
Fish and chips \$12	Battered fish served with potato fries
Eggs \$12	Poached eggs on toasted rye bread, served with spinach and mushrooms
Vegetarian risotto \$14	Mushroom and blue cheese risotto topped with rocket
Caesar salad \$8	Lettuce, parmesan, tomatoes with garlic croutons with homemade Caesar dressing
Tomato salad \$8	Tomato, mozzarella cheese and fresh basil leaves
Drink menu	
Drip coffee \$3	
Sparkling mineral water \$4	
Orange juice \$3	
Ginger Ale \$3	
Desert menu	
Tiramisu \$4	
Fresh fruit and ice cream \$4	

Appendix 15 Multi-word expression cloze test with answers and screenshot

Each question appeared on a new Moodle quiz page as shown in the screen shot here

Question 1
Not yet answered
Marked out of 4.00
Flag question
Edit question

H___ d___ y___ s___ your name?
お名前の綴りをお伺いできますか?

Next page

QUIZ NAVIGATION

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25
26	27	28	29	30

Finish attempt ...

Text used for the cloze test questions with answers shown in parentheses

1 Hotel H___ d___ y___ s___ your name? (How do you spell)

名前の綴りをお願いできますでしょうか。

2 Hotel T_____ y_____ v_____ m_____ (Thank you very much)

どうもありがとうございました。

3 Hotel How many are in your party? (How many are in)

グループは何名様でしょうか。

4 Hotel I'm sorry, c_____ y_____ r_____ t_____ please?

(could you repeat that)

すみません。もう一度おっしゃっていただけますか

5 Hotel H___ w___ y___ b___ paying? (How will you be)

お支払はいかがなさいますか。

6 Hotel T___ w___ b___ f____. (That will be fine)

大丈夫です。

7 Hotel I___ i___ a___ e___ \$15 each (It is an extra)

お一人様 15ドルの別料金になります。

8 Hotel W___ y___ l___ t___ add it to your reservation?

(Would you like to) ご予約に追加いたしますか。

9 Hotel T___ d___ a___ i___ your room (The details are in)

お部屋に説明書がございます。

10 Hotel L___ m___ t___ a___ it. (Let me think about)

検討させてください。

11 Café W___ w___ y___ l___ Mike? (What would you like)

マイクさんは何にしますか。

12 Café W___ h___ a___ a___ hour for lunch (We have about an)

お昼休みは一時間程度あります。

13 Café Let's see what they h___ o___ t___ m___. (have on the menu)

メニューに何があるか、見てみましょう。

14 Café The salmon looks good, b___ s___ d___ t___ omelette

(but so does the)

サケが美味しそうですけ、オムレツもおいしそうですね

15 Café W___ a___ y___ g___ to get? (What are you going)

何を頼みますか。

16 Café I t___ I w___ have the chicken. (I think I will)

チキンを頼もうかなと思っています

17 Café OK, I w___ h___ t___ salmon. (I will have the)

じゃ、私はサケにします。

18 Café A___ y___ g___ t___ get a coffee? (are you going to)

コーヒーを注文しますか。

19 Café Are you going t___ g___ a d___? (to get a drink)

飲み物を注文しますか。

20 Café I u___ o___ h___ time for a quick lunch (I usually only have)

普段は時間がなくて、簡単なお昼しか取れないです。

21 Directions H___ d___ I g___ to the train station? (How do I get)

駅はどうやって行きますか

22 Directions C___ y___ t___ m___ how to get to the post-office?

(Can you tell me)

郵便局への道を教えてくださいませんか。

23 Directions W___ y___ g___ t___ the traffic lights, turn left

(When you get to)

信号に着いたら左にまがってください

24 Directions T___ l___ a___ t___ traffic lights (turn left at the)

信号で左折する

25 Directions Walk along Green street u___ y___ g___ t___ the Square

(until you get to)

広場までグリーンストリートに沿って歩いて

26 Directions Y___ s___ s___ t___ Museum on the right. (You should see the)

博物館は右に見えるはずです

27 Directions The Museum is on the r___ h___ s___ o___ the Square.

(right hand side of)

博物館は広場の右手にあります。

28 Directions So, I turn left at the traffic lights then go up Green Street

u___ I g___ t___ the Square. (until I get to)

では、信号で左折して、そして、広場までグリーンストリートを直進して通ります。

29 Directions The Museum i___ o___ t___ r___ side of the Square


(is on the right)

博物館は広場の右手にあります。

30 Directions No worries, h___ a g___ d___! (have a good day)

いいえ、よい一日を！

Appendix 16 Activity feedback survey format example with translation



Please indicate to what extent you agree with the following statements
 下の文書はあなたの意見にどの程度当てはまりますか。

学生番号

下の文書はあなたの意見にどの程度当てはまりますか

	全く当てはまらない	当てはまらない	どちらでもない	当てはまる	非常に当てはまる
Marking pausesは簡単すぎでした。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Marking pausesは難しすぎでした。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Marking pausesは英語練習に役に立った。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

コメント

>>

Student number:

Please indicate to what extent you agree with the following statements

	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
<u>Marking pauses</u> was too easy					
<u>Marking pauses</u> was too difficult					
<u>Marking pauses</u> was useful for practising English					
Comment					

Appendix 17 Dictogloss worksheet from the café unit in Study One

Listen to the audio and take notes, then try to write the conversation with your partner. The words do not have to be the same as the audio, but the meaning should be similar

P: _____ We have about an _____.
_____.

J: _____. I usually only have
_____.

P: _____ have on the menu. What would you like _____?

J: _____, but so does the _____. What are you going
_____?

P: I think I'll have _____.

J: _____, I'll have the _____.

P: _____?

J: _____, _____ are you going to get a drink?

P: _____, _____, _____?

J: _____, _____.

P: _____, _____!

Appendix 18 List of multi-word expressions (MWEs) used in Study One

The spoken MWEs that were counted are shown with underlining

Pre-intervention control group

Word count	Spoken MWE	Target MWE	Unit	Form
2	Do you x 12	How do you spell	Hotel	question
2	Thank you	Thank you very much	Hotel	response
2	You like x 3	Would you like to	Hotel	question
2	<u>I think</u> we need	I think I will	Café	information
2	<u>a drink</u>	to get a drink	Café	question
4	Would you like to	Would you like to	Hotel	question

Post-intervention control group

2	Do you x 14	How do you spell	Hotel	question
2	<u>I will</u> order it	I think I will	Café	information
2	did <u>you like</u>	Would you like to	Hotel	question
2	Do <u>you like</u> x 3 (this would be counted as 3 in Study Two&3)	Would you like to	Hotel	question
2	<u>I think</u> fish	I think I will	Café	information
2	The menu	have on the menu	Café	information
2	<u>Are you</u> okay	what are you going	Café	question
2	<u>I think</u> that	I think I will	Café	information
2	<u>I think</u> it's good	I think I will	Café	information
2	<u>are you</u> sure	what are you going	Café	question
3	<u>I think</u> yes <u>I</u> want	I think I will	Café	information

Pre-intervention experimental group

2	Do <u>you like</u> x 9 (this would be counted as 3 in Study Two&3)	Would you like to	Hotel	question
2	Do you x 24	How do you spell	Hotel	question
2	are you	what are you going	Café	question
2	<u>I think</u> that one x 2	I think I will	Café	information
2	<u>It is</u> x 2	It is an extra	Hotel	information
2	<u>We have</u> x4	We have about an	Café	information
2	<u>To get</u> out	To get a drink	Café	question
2	I think	I think I will	Café	information
2	we have	We have about an	Café	information
3	what are you	what are you going	Café	question

Post-intervention experimental group

2	Do <u>you like</u> x 9 (this would be counted as 3 in Study Two&3)	Would you like to	Hotel	question
2	I have <u>a drink</u>	To get a drink	Café	question
2	May <u>I have</u>	I will have the	Café	information
2	Do you x 28	How do you spell	Hotel	question
2	<u>we have</u> one hour x 4	We have about an	Café	information
2	<u>Are you</u> hungry	what are you going	Café	question
2	<u>I will</u> order	I think I will	Café	information
2	<u>I think</u> x 2	I think I will	Café	information
2	Thank you x 7	Thank you very much	Hotel	response
2	<u>Can you</u> eat	Can you tell me	Directions	question
2	Could you	Could you repeat that	Hotel	question
2	<u>We have</u> to order x 2	We have about an	Café	information

2	<u>Are you</u> okay	what are you going	Café	question
2	Do you <u>have a</u>	Have a good day	Directions	greeting
3	<u>How do you</u> pay	How do you spell	Hotel	question
3	would you like	would you like to	Hotel	question
3	We have about	We have about an	Café	information
3	<u>We have an</u> hour	We have about an	Café	information
4	Have a good day	Have a good day	Directions	greeting
4	would you like to x 2	would you like to	Hotel	question
5	what would you like to x 2	What would you like/would you like to	Hotel	question

Appendix 19 List of extended multi-word expressions used in Study One

(words counted as extended are shown in parentheses)

Used in pre-intervention dialogue

(do) you like

I think (we need)

I think (that one)

(to) get out

I think (we have)

Used in post-intervention dialogue

I will (order it)

Are you (hungry)

(did) you like

I will (order)

I think (fish)

How do you (pay)

are you (okay)

(and) would you like to

I think (yes) I (want)

(can) you eat

I think (that)

We have (to order)

I think (it's good)

We have (two chicken)

are you (sure)

I think (probably)

(do) you like

We have about (hour)

(I have) a drink

We have an (hour)

(may) I have

(Do you) have a (dessert)

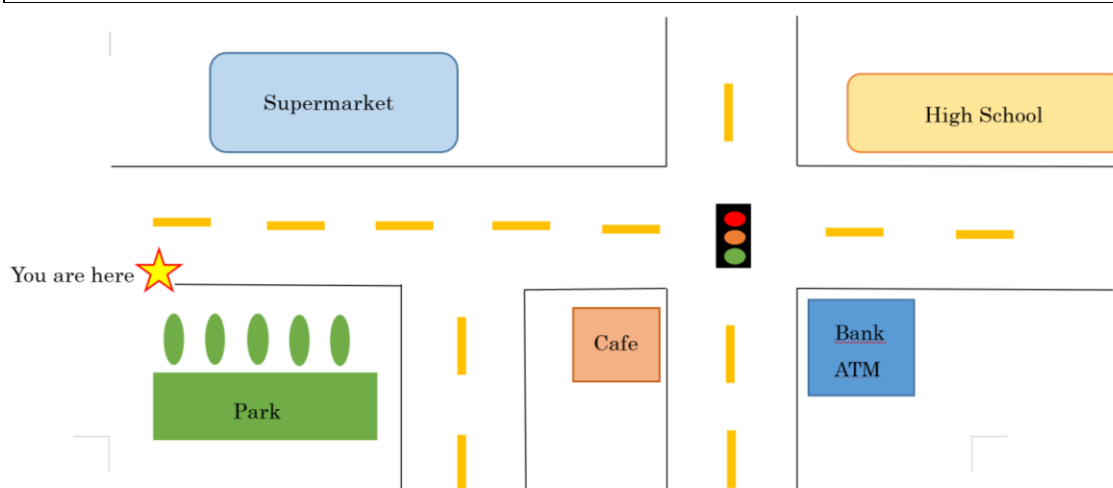
We have (one hour)

Appendix 20 Adapted pre- post- intervention role-play instructions


Role-play the following scenario with your partner

下記の場면을相手と一緒にロールプレイする。

Person #1	Person #2
1. say your name (<i>first person</i>) 名前と <i>first person</i> を言う	2. say your name (<i>second person</i>) 名前と <i>second person</i> を言う
3. Ask your partner for directions to the nearest ATM. もよりのATMはどこにあるかを聞く。	4. Give directions to ATM using the map 地図を使って道案内をする
5. Invite your partner to go to lunch at the café 相手をランチに誘う	6. Accept and go to café together 受けて、一緒にカフェへ行く
7. Look at menu and talk about what you want to eat メニューを見て何を頼むか話しあって決める	
8. Pretend to be the waiter and ask for order 店員の役をして注文を聞く	9. Tell the waiter what you want 自分の注文を伝える
10. Take the order 注文を受け取る	11. Pretend to be the waiter and ask for order 店員の役をして注文を聞く
12. Tell the waiter what you want 自分の注文を伝える	13. Take the order 注文を受け取る
14. Your food arrives, say <i>thank you</i> and eat, then finish. 食事が届く、 <i>ありがとう</i> と言って、食べて、終わる。	



Appendix 21 Adjusted activity feedback survey example



Please indicate to what extent you agree with the following statements
 下の文書はあなたの意見にどの程度当てはまりますか。

学生番号

下の文書はあなたの意見にどの程度当てはまりますか

	非常に当ては まらない	当てはまらな い	少し当てはま らない	少し当てはま る	当てはまる	非常に当ては まる
フレーズ説明は社会的に難しい でした。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
フレーズ説明は技術的に難しい でした。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
フレーズ説明は言語的に難しい でした。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
フレーズ説明は英語練習に役に 立った。	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
フレーズ説明は楽しかった	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

コメント

Survey Powered By Qualtrics

English translation (the activity name is underlined)

Please indicate to what extent the following statements are true

1. Phrase instruction was socially difficult
2. Phrase instruction was technically difficult
3. Phrase instruction was linguistically difficult
4. Phrase instruction was useful for practising English
5. Phrase instruction was enjoyable

Comment

Appendix 22 Technical support confidentiality agreement



Transcribing Confidentiality Agreement

Project Title: Developing fluency with multi-word expressions

Principal Investigator: Haidee Thomson

I, _____, agree to ensure that the audio files I transcribe will remain confidential to Haidee Thomson and myself.

I agree to take the following precautions:

1. I will ensure that no other person hears the recording.
2. I will ensure that no other person has access to my computer/device.
3. I will delete the files from my computer/device once the transcription has been completed.
4. I will not discuss any aspect of the recording with anyone except Haidee Thomson.

Signature: _____

Date: _____

Appendix 23 Fluency reflection worksheet

Name _____

Student number _____

Listen to your role-play recording and draw a circle on the continuum to show your conversation performance.

録音した会話を聞いて、自分の会話結果を下記の連続体に当てるところで丸を描いてください。

Restaurant

Long pauses

Flowing conversation

長い休憩

流れている会話

--	--	--	--	--	--	--	--	--	--	--

Difficult to hear

Clear pronunciation

聞きにくい

発音ははっきり言っている

--	--	--	--	--	--	--	--	--	--	--

Word by word

Phrase by phrase

言葉ごと休憩

フレーズごと休憩

--	--	--	--	--	--	--	--	--	--	--

Directions

Long pauses

Flowing conversation

長い休憩

流れている会話

--	--	--	--	--	--	--	--	--	--	--

Difficult to hear

Clear pronunciation

聞きにくい

発音ははっきり言っている

--	--	--	--	--	--	--	--	--	--	--

Word by word

Phrase by phrase

言葉ごと休憩

フレーズごと休憩

--	--	--	--	--	--	--	--	--	--	--

Hotel

Long pauses

Flowing conversation

長い休憩

流れている会話

--	--	--	--	--	--	--	--	--	--	--

Difficult to hear

Clear pronunciation

聞きにくい

発音ははっきり言っている

--	--	--	--	--	--	--	--	--	--	--

Word by word

Phrase by phrase

言葉ごと休憩

フレーズごと休憩

--	--	--	--	--	--	--	--	--	--	--

Appendix 24 Adjusted dictogloss example from café unit

Dictogloss

- 1) Listen to the audio, no writing.
- 2) Listen again and take notes (words that will help you re-write later x3)
- 3) Work with your group to re-write text (different words are okay)

1 Pam	We have about an
2 John	I usually only have
3 Pam	have on the menu. What would you like
4 John	but so does the what are you going
5 Pam	I think I'll
6 John	I'll have the
7 Pam	
8 John	are you going to get a drink?
9 Pam	
10 John	
11 Pam	

Appendix 25 List of multi-word expressions (MWEs) used in Study Two

Pre-intervention use of multi-word expressions in the dialogue recording

word count	Spoken MWE	Target MWE	Unit	Form
0	Turn right x 2	turn left at the	Directions	instruction
2	right side	right hand side of	Directions	information
2	<u>Can you</u> take x 2	can you tell me	Directions	question
2	lunch <u>at the</u> café	turn left at the	Directions	instruction
2	<u>What do you</u> want	what would you like	Café	question
2	Thank you x13	Thank you very much	Hotel	response
3	<u>What do you like</u>	what would you like	Café	question
3	<u>Can you help me</u>	can you tell me	Directions	question
4	What do you want	what would you like	Café	question
4	How can I go to	How do I get	Directions	question

Post-intervention use of multi-word expressions in the dialogue recording

word count	Spoken MWE	Target MWE	Unit	Form
2	<u>Thank you</u> x 8	Thank you very much	Hotel	response
2	<u>are you</u> x 3	what are you going	Café	question
2	<u>I will</u> choose /order x 2	I think I will	Café	information
2	<u>What do you</u> want x 3	what would you like	Café	question
2	<u>Right hand</u>	right hand side of	Directions	information
2	lunch <u>at the</u> café x2	turn left at the	Directions	instruction
2	<u>You can see</u> x2	you should see the	Directions	instruction
3	<u>You see the</u>	you should see the	Directions	instruction
3	<u>How</u> can <u>I get</u> to	How do I get	Directions	question
3	<u>How I get</u> to	How do I get	Directions	question
4	Can you tell me	Can you tell me	Directions	question
4	I think I'll	I think I will	Café	information

Appendix 26 Hotel booking form realia from Study Three

Hotel booking form

Number of people	
Room type	
Number of nights	
Name	
Payment	
Breakfast	
WIFI	
Room number	

Appendix 27 List of multi-word expressions (MWEs) used in Study Three

Count	<i>Pre-control MWEs used</i>	<i>Target</i>
2	Thank you	Thank you very much
4	Thank you very much	Thank you very much
2	<u>You can see</u>	You should see the
2	You should	You should see the
3	<u>You will see the</u>	You should see the
3	<u>You can see the</u>	You should see the
2	We have	We have about an
4	We have a good	we have about an & have a good day
4	<u>Could you tell me</u>	Could you repeat that & Can you tell me
3	On the right	Is on the right
2	<u>At the</u> café	Turn left at the
2	Do you	How do you spell
3	Do you get	how do you spell & when you get to
2	Here <u>it is</u>	It is an extra
2	Are you	What are you going
2	<u>I wanna have</u>	I will have the
2	I m <u>gonna</u> have	Are you going to
2	You want <u>to * drink</u>	To get a drink
3	<u>Would you like</u> some	Would you like to
4	Would you like to	Would you like to
5	What would you like to	would you like to & what would you like

Count	<i>Post-control MWEs used</i>	<i>Target</i>
2	Thank you	Thank you very much
2	Your <u>right hand</u>	Right hand side of
2	See <u>the right</u> side	Is on the right
2	<u>You can see</u>	You should see the
3	<u>You look see the</u>	You should see the

3	<u>You can see the</u>	You should see the
4	When you see the	when you get to & you should see the
2	<u>Do you</u> know	How do you spell
3	Would you like	Would you like to
4	Would you like to	Would you like to
5	What would you like to	would you like to & what would you like
2	Something <u>to * drink</u>	To get a drink
2	I ll take	I think I will & I will have the
2	Will be	That will be fine
2	What <u>will you</u> take	How will you be
2	Look <u>at the</u> map	Turn left at the
2	<u>It is</u> very delicious	It is an extra
2	The menu	Have on the menu
4	<u>Could you tell me</u>	Could you repeat that & Can you tell me

<i>Count</i>	<i>Pre-experimental MWEs used</i>	<i>Target</i>
2	Do you	How do you spell
3	Do you like	how do you spell & would you like to
2	I get	How do I get
2	Can you	Can you tell me
2	<u>At the</u>	Turn left at the
2	On the	Is on the right
2	<u>On your right</u>	Is on the right
2	<u>it is</u>	It is an extra
2	<u>You can see</u>	You should see the
2	<u>You can see</u>	You should see the
2	Thank you	Thank you very much
2	Are you	What are you going
2	The menu	Have on the menu

2	We have	We have about an
2	Will you	How will you be
3	<u>I * have the</u>	I will have the
2	How many	How many are in
4	Would you tell me	would you like to & could you tell me

<i>Count</i>	<i>Post-experimental MWEs used</i>	<i>Target</i>
2	Do you	How do you spell
3	Do you like	how do you spell & would you like to
2	would you	What would you like
4	Would you tell me	would you like to & can you tell me
4	would you like to * drink	would you like to & to get a drink
4	Would you like to	Would you like to
4	What would you like	What would you like
5	<u>Would you like</u> anything to <u>*drink</u>	would you like to & to get a drink
5	Would you like to get	would you like to & to get a drink
5	What would you like to	would you like to & what would you like
2	Should see	You should see the
3	You should see	You should see the
2	See the	You should see the
2	<u>You can see</u>	You should see the
3	<u>You can see the</u>	You should see the
5	<u>You can see you get to</u>	you should see the & when you get to
2	On the	Is on the right
2	the right	Is on the right
3	On the right	Is on the right
5	The right hand side on the	right hand side of & is on the right
2	Right hand	Right hand side of

2	Left <u>hand side</u>	Right hand side of
2	Turn left	Turn left at the
2	At the	Turn left at the
3	Look <u>at the menu</u>	turn left at the & have on the menu
4	Thank you very much	Thank you very much
2	I will	I think I will
2	I ll	I think I will
3	I think I	I think I will
2	Will you	How will you be
2	Are you	What are you going
2	I get	Until I get to
2	<u>to * drink</u>	To get a drink
3	How to get	how do I get & to get a drink
4	How do I get	How do I get
4	<u>How</u> can I <u>get to</u>	how do I get to & until I get to
4	How I get to	how do I get & until I get to
7	Can you tell me how to get	can you tell me & how do I get & to get a drink
4	When you get to	When you get to
4	Could you tell me	could you repeat that& can you tell me