Investigating the impact of learner codeswitching on L2 oral fluency in task-based activities: The case of EFL primary school classrooms in Cyprus.

Maria Vrikki

Thesis submitted to the University of Oxford for the Degree of Doctor of Philosophy in Education

Hilary Term 2013

Kellogg College

Abstract

The potentially beneficial role of classroom codeswitching, or the use of the first language (L1) in foreign language (FL) classroom settings, is gradually becoming acknowledged in the field of second language acquisition (SLA) research. However, researchers call for the construction of a framework, which indicates when this use is beneficial for language learning and when it is not. In an attempt to contribute to the construction of this framework, the present study investigates whether codeswitching can be used as a tool within task-based learning settings for the development of second language (L2) oral fluency. It is hypothesised that by allowing learners to codeswitch during task completion, their willingness to communicate (WTC) is enhanced because the function of that switching is likely to be mainly the metalanguage needed to complete the task.

Previous research has also suggested that task repetition might lead to greater fluency. However, both teachers and learners may be sceptical of the value of repetition without some form of feedback on the first task attempt. This study sought to explore therefore the value of task repetition with feedback (TR+). By repeating the task with feedback that recycles metalanguage into the L2, it is hypothesised that learners will learn to move to a state of less reliance on their L1, while simultaneously achieving the overall aim of tasks, which is effective L2 communication. With increased WTC and L2 metalanguage, extensive L2 oral practice will facilitate the proceduralisation processes needed for fluency development. In short, the pedagogical package of TR+ on recycled language is tested in the present study as a potential contributor to oral fluency.

The thesis begins by relating these themes with the context of Cyprus through teacher interviews. It becomes evident through these interviews that the activities taking place in this context are not *tasks* in the sense researchers intended. Following the setting of the context, the interactions of 75 primary school learners of English (11-12-year-olds) practising TR+ are analysed qualitatively. This analysis determines whether the package can lead to enriched output on the second attempt. In addition, there is a quasi-experimental aspect to the study. The students were allocated in three groups, each testing a different package. The *codeswitching group* was allowed to switch to Greek while completing the tasks and had their L1 metalanguage recycled into the L2 when they repeated the tasks. The *English-only group* completed the tasks strictly under L2 conditions and repeated them with feedback on accuracy. The *comparison group* completed the tasks once with no language instructions. Oral production tests, used as pre- and post-tests, partly support the hypothesis by suggesting that WTC is enhanced with the incorporation of codeswitching, but no evidence supports fluency development. Nevertheless, when comparing TR+ with no task repetition, the data indicate that TR+ leads to greater fluency.

It is suggested that a larger and longer intervention would have allowed more time for fluency to be developed when codeswitching was incorporated. As for task-based learning, it is suggested that TR+ is a more viable way to move forward in real classroom contexts, particularly those with young learners. Furthermore, the results of the present study indicate that this package works better with learners of a certain proficiency level.

This thesis is dedicated to my mother, Dr. Dimitra Karoulla-Vrikki

Acknowledgements

I am highly indebted to my supervisor, Professor Ernesto Macaro. Through his constant guidance and support, as well as his invaluable insights and expertise, he has immensely contributed to the successful completion of this thesis.

I would also like to express my gratitude to Dr. Robert Vanderplank, Dr. Victoria Murphy, Dr. Catherine Walter, Professor Kathy Sylva and Dr. Parvaneh Tavakoli. Their constructive feedback, at various stages of the thesis, such as the Final Viva, the Confirmation of Status and the Transfer of Status, has been decisive in improving and shaping the final version of the thesis. Special thanks go to the library staff for their eagerness to help in locating sources for me.

I am very grateful to the children who participated in the study and made this experience an unforgettable one; to their teachers, who welcomed me in their classrooms and helped me in any way they could; to the teachers who participated in the interview process; and to the teachers who allowed me access to their schools for months so that I could carry out my study. Without them, this study would not have been possible.

Finally, I am forever indebted to my parents, George and Dimitra, for the inexhaustible love, care and encouragement expressed in so many ways during all my studies. Without having such great role models, I would not be the person I am today. Thank you to my younger brother, Andreas, for his love and support. And last, but not least, thanks go to Pantelis for having faith in me and for standing by my side throughout these years.

Table of Contents

CHAPTER 1: Introduction

1.1. Aims of the study....1
1.2. Context of the study....2
1.3. Overview of the thesis....5

CHAPTER 2: Literature Review

2.1. Introduction....7
2.2. Task-based learning....7
 2.2.1. Defining tasks....10
 2.2.2. Teachers' opinions on tasks....12
 2.2.3. Empirical studies on task-based learning....16
 2.2.4. Task repetition....19
 2.2.5. Conclusions on task-based learning....28
2.3. Codeswitching....29
 2.3.1. Defining codeswitching....29
 2.3.2. Issue of controversy....31
 2.3.3. Empirical studies....36
  2.3.3.1. Studies on teacher codeswitching....36
   i. Descriptive studies....36
   ii. Experimental studies....40
  2.3.3.2. Studies on teachers' and students' perceptions....42
  2.3.3.3. Studies on learner codeswitching....46
   i. Vygotsky's Sociocultural Theory (SCT)....47
   ii. Studies with non-adult participants....48
   iii. Studies with adult participants....53
   iv. Conclusions on learner codeswitching studies....60
 2.3.4 Conclusions on codeswitching....62
2.4. Oral fluency....64
 2.4.1 Defining fluency....64
 2.4.2. Measuring fluency....64
 2.4.3. Fluency development....67
 2.4.4. Conclusions on fluency....71
2.5. Willingness to Communicate....72
 2.5.1. MacIntyre's models....72
 2.5.2. Measuring WTC....74
 2.5.3. Conclusions on WTC....79
2.6. Hypotheses....80
2.7. Chapter summary....83

CHAPTER 3: METHODOLOGY

3.1. Introduction....85
3.2. Research Questions....85
3.3. Research design....88
3.4. Population - Sample....91
3.5. The intervention....93
3.5.1. Tasks....94
3.5.2. Conditions in the experiment....96
i. Codeswitching group (Experimental)....96
ii. English-only group (Experimental)....98
iii. Comparison group....99
3.5.3. Oral production tests....100
3.5.4. Fidelity to conditions....104
3.6. Ethical considerations....108
3.7. Pilot study....109
3.7.1. Piloting the experimental conditions....109
i. English-only condition....109
ii. Codeswitching condition....110
3.7.2. Piloting tasks....111
3.7.3. Piloting the oral production test....113
3.8. Matching participants....113
3.8.1. Second Language Proficiency....113
3.8.2. Productive Vocabulary....116
3.8.3. Socio-economic status (SES)....119
3.8.4. Prior contact with the English language....120
3.9. Other confounding variables....121
3.9.1. Researcher's protocol....121
3.9.2. Recording Factor....122
3.10. Teacher Interviews....123
3.10.1. Sampling procedure – Ethical considerations....123
3.10.2. The Interviews....124
3.10.3. Interview Schedule....124
3.11. Data Analysis....125
3.11.1. Analysis for Research Question 1....125
3.11.2. Analysis for Research Question 2....125
3.11.3. Analysis for Research Question 3....126
3.11.4. Analysis for Research Question 4....128
i. Fluency variables....128
ii. Content variables....131
iii. Accuracy variables....133
3.11.5. Analysis for Research Question 5....135
3.12. Limitations....137
3.13. Chapter summary....139

CHAPTER 4: Qualitative Analysis of Teacher Interviews

4.1. Introduction..141
4.2. Task implementation...141
4.2.1. The value of tasks..141
4.2.2. Factors affecting task implementation...145
4.2.3. Experiences of teaching...149
4.2.4. The new curriculum...152
4.2.5. Conclusions on task implementation...153
4.3. Task repetition...155
4.4. Metalanguage...158
4.5. Combinations of pairs in terms of proficiency...163
4.5.1 Teachers' views on mixed-proficiency and same-proficiency pairs................163
4.5.2 Scaffolding..167
4.5.3 Interlocutor familiarity..169
4.5.4 Conclusions on pair work..172
4.6. Chapter summary..173

CHAPTER 5: Qualitative analysis of student interactions

5.1. Introduction..176
5.2. Data and sample..177
5.3. Qualitative analysis results...178
5.3.1. Task language...179
i. Examples of richer task language on S2..179
ii. Examples of poorer task language on S2...183
5.3.2. L2 Metalanguage...188
i. L2 metalanguage of CS group...189
ii. L2 metalanguage of EO group..198
5.4. Conclusions...204
5.4.1. Scaffolding..204
5.4.2. L1 metalanguage...205
5.4.3. Effects of accuracy feedback..206
5.4.4. Enriched output...206
i. Task language..206
ii. L2 metalanguage..207
5.5. Chapter summary..209

CHAPTER 6: Quantitative Results

6.1. Introduction....211
6.2. Possible confounding variables....211
6.2.1 Second Language Proficiency....211
6.2.2 Second Language Productive Vocabulary....213
6.2.3 Socio-economic status....214
6.2.4 Prior L2 contact....217
6.2.5 Matching Process Results....218
6.3. RQ3: a) Does task repetition plus feedback lead to more willingness to communicate than no task repetition?
b) Does task repetition with codeswitching plus lead to more willingness to communicate than task repetition with English-only plus? 219
6.3.1. Speaking time....219
6.3.2. L2 Syllables....222
6.3.3 Response to RQ3....224
i. Task repetition plus vs. no task repetition....225
ii. Task repetition with CS plus vs. task repetition with EO plus....225
6.4. RQ4: a) Does task repetition plus feedback lead to more fluency than no task repetition?
b) Does task repetition with codeswitching plus lead to more fluency than task repetition with English-only plus, with no detrimental effects on accuracy?....226
6.4.1 Fluency Variables....226
i. Speech Rate....226
ii. Mean Length of Runs....228
iii. Pause Rate....231
iv. Repetitions....233
v. Reformulations....235
vi. Incomplete words....238
6.4.2 Content variables....240
i. Verb Ratio....240
ii. L1 syllables....242
iii. Idea units....245
6.4.3 Accuracy variables....247
i. Percentage of error-free clauses....247
ii. Percentage of correct verb use....249
6.4.4. Response to RQ4....251
i. Task repetition plus vs. no task repetition....254
ii. Task repetition with CS plus vs. Task repetition with EO plus....255
6.5. RQ5: Do higher proficiency students benefit more from task repetition *plus* feedback than do lower proficiency students? 256
6.5.1. Fluency variables....256
i. Speech Rate....256
ii. Mean Length of Runs....258
iii. Pause Rate....260
iv. Repetitions....262

v. Reformulations..264
vi. Incomplete words..265
6.5.2. Content variables..267
i. Verb Ratio..267
ii. L1 syllables..269
iii. Idea units..271
6.5.3 Accuracy variables..273
i. Percentage of error-free clauses..273
ii. Percentage of correct verb use..274
6.5.4 Response to RQ5..276
6.6. Chapter Summary..280

CHAPTER 7: Discussion

7.1. Introduction..282
7.2. Discussion on RQ1..282
7.2.1. Task implementation..282
7.2.2. Task repetition..287
7.2.3. Metalanguage..289
7.2.4. Proficiency pairing..291
7.2.5. Conclusions on RQ1..293
7.3. Discussion on RQ2..294
7.3.1. Task language..294
7.3.2. L2 metalanguage..295
7.4. Discussion on RQ3..296
7.5. Discussion on RQ4..298
7.6. Discussion on RQ5..302
7.7. Chapter summary..303

CHAPTER 8: Conclusion

8.1. Introduction..306
8.2. Conclusions and pedagogical implications..306
8.3. Limitations..308
8.4. Contribution to the field..310
8.5. Recommendations for future work..311

References..313

List of Tables

Table 2.2.1 Graphic representation of Ahmadian's (2011) intervention design....................25
Table 2.3.1 Studies calculating the amount of teacher codeswitching..................................37
Table 2.3.2 Studies describing the functions of teacher codeswitching................................39
Table 3.3.1 Independent and dependent variables...91
Table 3.4.1 Number of participants..93
Table 3.5.1 Graphic representation of design..99
Table 3.5.2 L1/L2 distribution of two teachers..107
Table 3.11.1 Correlations for intra-rater reliability: WTC measures..................................128
Table 3.11.2 Description of fluency variables...130
Table 3.11.3 Correlations for intra-rater reliability: Fluency measures...............................130
Table 3.11.4 Description of content variables...132
Table 3.11.5 Correlations for intra-rater & inter-rater reliability: Content measures...........133
Table 3.11.6 Correlations for intra-rater reliability: Accuracy measures.............................134
Table 3.11.7 Summary of research questions & method of analysis...................................136
Table 6.2.1 Descriptive statistics: L2 Proficiency (C-test)..212
Table 6.2.2 Descriptive Statistics: Productive Vocabulary Scores.......................................213
Table 6.2.3 Descriptive Statistics: SES scores...215
Table 6.2.4 Matching process results...218
Table 6.3.1 Descriptive Statistics: Speaking time..221
Table 6.3.2 Descriptive Statistics: L2 syllables...223
Table 6.3.3 Results of WTC measures...225
Table 6.4.1 Descriptive Statistics: Speech Rate...227
Table 6.4.2 Descriptive Statistics: Mean Length of Runs scores...229
Table 6.4.3 Descriptive Statistics: Pause Rate...232
Table 6.4.4 Descriptive Statistics: Repetitions..234
Table 6.4.5 Descriptive Statistics: Reformulations..236
Table 6.4.6 Descriptive Statistics: Incomplete words..239
Table 6.4.7 Descriptive Statistics: Verb Ratio...241
Table 6.4.8 Descriptive Statistics: L1 Syllables..243
Table 6.4.9 Descriptive Statistics: Idea units...246
Table 6.4.10 Descriptive Statistics: Percentage of error-free clauses..................................248
Table 6.4.11 Descriptive Statistics: Percentage of correct verb use....................................250
Table 6.4.12 Results of temporal variables..252
Table 6.4.13 Results of dysfluency markers..252
Table 6.4.14 Results of content variables..253
Table 6.4.15 Results of accuracy variables..254
Table 6.5.1 Descriptive Statistics: Speech Rate (Proficiency groups).................................257
Table 6.5.2 Descriptive Statistics: MLR (Proficiency groups)..259
Table 6.5.3 Descriptive Statistics: Pause Rate (Proficiency groups)...................................261
Table 6.5.4 Descriptive Statistics: Repetitions (Proficiency groups)...................................263
Table 6.5.5 Descriptive Statistics: Reformulations (Proficiency groups).............................265
Table 6.5.6 Descriptive Statistics: Incomplete words (Proficiency groups).........................266
Table 6.5.7 Descriptive Statistics: Verb Ratio (Proficiency groups)...................................268

Table 6.5.8 Descriptive Statistics: L1 Syllables (Proficiency groups).................................270
Table 6.5.9 Descriptive statistics: Idea units (Proficiency groups).....................................272
Table 6.5.10 Descriptive Statistics: Percentage of error-free clauses (Proficiency groups).274
Table 6.5.11 Descriptive Statistics: Percentage of correct verb use (Proficiency groups)..275
Table 6.5.12 Results of temporal variables of proficiency groups......................................277
Table 6.5.13 Results of dysfluency markers of proficiency groups....................................277
Table 6.5.14 Results of content variables of proficiency groups..278
Table 6.5.15 Results of accuracy variables of proficiency groups......................................279

List of figures

Figure 2.2.1 The Triadic Componential Framework (Robinson, Cadierno & Shirai, 2009: 535) 17
Figure 2.2.2 Gass *et al.* (1999) intervention design (p. 556)..21
Figure 2.2.3 Bygate's (2001) intervention design...23
Figure 2.4.1 Levelt's Model of Speech Production (Levelt, 1989: 9)..................................68
Figure 2.4.2 Anderson's Adaptive Control of Thought Model (J.R. Anderson, 1983: 19)..69
Figure 2.5.1 MacIntyre et al. (1998) pyramid model of WTC...73
Figure 2.6.1 Graphic representation of hypotheses...82
Figure 3.5.1 Percentage of L1 used by the two experimental groups on four weeks............105
Figure 6.2.1 Boxplots: L2 Proficiency scores of the three groups (C-test scores)................212
Figure 6.2.2 Boxplots: Productive Vocabulary Scores of the three groups (Translation test)...213
Figure 6.2.3 Boxplots: SES scores of the three groups..215
Figure 6.2.4 Histogram: SES classification of the three groups...216
Figure 6.2.5 Histograms: Prior L2 contact classification...217
Figure 6.3.1 Line graph: Mean performance of speaking time of the three groups across time..220
Figure 6.3.2 Line graph: Mean performance of the amount of L2 syllables of the three groups across time...223
Figure 6.4.1 Line graph: Line graph: Mean performance of speech rate of the three groups across time...227
Figure 6.4.2 Line graph: Mean performance of MLR of the three groups across time..229
Figure 6.4.3 Line graph: Mean performance of pause rate of the three groups across time..232
Figure 6.4.4 Line graph: Mean performance of the amount of repetitions of the three groups across time...234
Figure 6.4.5 Line graph: Mean performance of the amount of reformulations of the three groups across time...236
Figure 6.4.6 Line graph: Mean performance of the amount of incomplete words of the three groups across time...238

Figure 6.4.7 Line graph: Mean performance of verb ratio of the three groups across time..........241
Figure 6.4.8 Line graph: Mean performance of the amount of L1 syllables of the three groups across time..........243
Figure 6.4.9 Line graph: Mean performance of the amount of idea units of the three groups across time..........245
Figure 6.4.10 Line graph: Mean performance of the percentage of error-free clauses of the three groups across time..........248
Figure 6.4.11 Line graph: Mean performance of the percentage of correct verbs of the three groups across time..........250
Figure 6.5.1 Line graph: Mean performance of speech rate of the two proficiency groups across time..........257
Figure 6.5.2 Line graph: Mean performance of MLR of the two proficiency groups across time..........259
Figure 6.5.3 Line graph: Mean performance of pause rate of the two proficiency groups across time..........260
Figure 6.5.4 Line graph: Mean performance of the amount of repetitions of the two proficiency groups across time..........262
Figure 6.5.5 Line graph: Mean performance of the amount of reformulations of the two proficiency groups across time..........264
Figure 6.5.6 Line graph: Mean performance of the amount of incomplete words of the two proficiency groups across time..........266
Figure 6.5.7 Line graph: Mean performance of verb ratio of the two proficiency groups across time..........268
Figure 6.5.8 Line graph: Mean performance of the amount of L1 syllables of the two proficiency groups across time..........270
Figure 6.5.9 Line graph: Mean performance of the amount of idea units of the two proficiency groups across time..........271
Figure 6.5.10 Line graph: Mean performance of the percentage of error-free clauses of the two proficiency groups across time..........273
Figure 6.5.11 Line graph: Mean performance of the percentage of correct verb use of the two proficiency groups across time..........275

List of appendices

Appendix 1: Week 1 Task..327
Appendix 2: Week 2 Task..329
Appendix 3: Week 3 Task..333
Appendix 4: Week 4 Task..334
Appendix 5: Week 5 Task..335
Appendix 6: Week 6 Task..336
Appendix 7: Week 7 Task..337
Appendix 8: Week 8 Task..338
Appendix 9: Fluency handout (Week 5) 340
Appendix 10: Accuracy handout (Week 6) 341
Appendix 11: Pre-test...342
Appendix 12: Post-test 1...343
Appendix 13: Post-test 2...344
Appendix 14: Excerpts from the transcript of Teacher A's lesson with codeswitching group – Week 4...345
Appendix 15: Excerpts from the transcript of Teacher B's lesson with comparison group – Week 4...347
Appendix 16: School letter..349
Appendix 17: School's consent form...351
Appendix 18: Teacher's letter...352
Appendix 19: Teacher's consent form...354
Appendix 20: Parents' information pamphlet (Codeswitching group)...............................355
Appendix 21: Parents' information pamphlet (English-only group)..................................356
Appendix 22: Parents' information pamphlet (Comparison group)...................................357
Appendix 23: Parents' consent form...358
Appendix 24: Students' consent slip...359
Appendix 25: Examples of L1 use under the English-only condition (Pilot study)............360
Appendix 26: Examples of L1 use under the Codeswitching condition (Pilot study)..........361
Appendix 27: C-test...362
Appendix 28: Productive Vocabulary Test...364
Appendix 29: SES Questionnaire..365
Appendix 30: Student Questionnaire..366
Appendix 31: Researcher's protocol...367
Appendix 32: Teacher letter (Interviews)...368
Appendix 33: Interview Schedule...369
Appendix 34: Additional examples of increased and reduced creativity of task language from session 1 to session 2 (Chapter 5)...370
Appendix 35: Normality distributions for possible confounding variables data.................373
Appendix 36: Distribution graphs for WTC measures...375
Appendix 37: Comparisons of results of parametric to non-parametric tests for WTC measures..378
Appendix 38: Distribution graphs for temporal measures of fluency................................379

Appendix 39: Distribution graphs for dysfluency measures of fluency..............................384
Appendix 40: Comparisons of results of parametric to non-parametric tests for all fluency measures..389
Appendix 41: Distribution graphs for content measures..391
Appendix 42: Distribution graphs for accuracy measures..396
Appendix 43: Comparisons of results of parametric to non-parametric tests for content and accuracy measures. 399
Appendix 44: Distribution graphs for temporal variables of fluency (for proficiency groups)..400
Appendix 45: Distribution graphs for dysfluency variables of fluency (for proficiency groups)..405
Appendix 46: Comparisons of results of parametric to non-parametric tests for fluency measures (for proficiency groups)...410
Appendix 47: Distribution graphs for content variables (for proficiency groups)................412
Appendix 48: Comparisons of results of parametric to non-parametric tests for content measures (for proficiency groups)...417
Appendix 49: Distribution graphs for accuracy variables (for proficiency groups).............418
Appendix 50: Comparisons of results of parametric to non-parametric tests for accuracy measures (for proficiency groups)...421

List of Abbreviations

CLT	Communicative Language Teaching
COM	Comparison
CS	Codeswitching
EFL	English as a Foreign Language
EL	Embedded Language
ELT	English Language Teaching
EO	English-only
FL	Foreign Language
HH	High proficiency-High proficiency
HL	High proficiency-Low proficiency
HP	High Proficiency
K-S test	Kormogorov-Smirnoff test
L1	First Language
L2	Second Language
LL	Low proficiency-Low proficiency
LP	Low proficiency
ML	Matrix Language
MLR	Mean Length of Runs
RQ	Research Question
S1	Session 1
S2	Session 2
SCT	Socio-cultural Theory
SLA	Second Language Acquisition
TL	Target Language
TR+	Task Repetition plus feedback
TTR	Type/Token Ratio

CHAPTER 1: Introduction

1.1. Aims of the study

From the early 1980s, the use of codeswitching, or else the use of the students' first language (L1), in foreign language (FL) classrooms was often ignored or frowned upon. The main reasons for this negative attitude held towards classroom codeswitching stem from the fact that this use contradicted the foundations of traditional language teaching methodologies (e.g. Direct method), as well as with influential input theories of second language acquisition (SLA) research (e.g. Krashen's *Comprehensible Input Hypothesis*).

Nevertheless, the large body of research investigating the codeswitching behaviour of both teachers and learners has revealed that codeswitching exists in various FL classroom settings and that it often has a facilitative role. Based on this finding, researchers call for the construction of a framework, which indicates when codeswitching contributes to language learning and when it is simply the easy option (Macaro, 2001). Such a framework will have major pedagogical implications, as teachers will become aware of when their switches to the L1 are acceptable and when they may possibly hinder language learning.

Recently, a few experimental studies have attempted to contribute to the construction of this framework by examining the potential effects of codeswitching on vocabulary acquisition (e.g. Tian, 2009; Lee, 2010). The present study aims to build on this new strand of research by investigating the potential beneficial role of codeswitching on oral fluency development within task-based learning settings. More specifically, it examines a pedagogical package involving codeswitching and then recycling this language into the L2, as a potential contributor to oral fluency. This innovative package of *codeswitching plus* is central to the present study.

The aims of the present thesis extend to the area of task-based learning, and more specifically, to task repetition. Previous research has suggested that task repetition could have positive effects on oral production, and particularly on fluency (e.g. Bygate, 2001; Ahmadian, 2011). However, as the results of studies in this area are not yet firm, the package of task repetition *plus* feedback is put forward in the present study. This package is considered to be more acceptable for language educators, as well as students, in real classroom contexts.

Both ideas proposed in the present study, namely the potential relationship between codeswitching and oral fluency, and the idea of task repetition *plus,* have not been studied before. In this sense, it is believed that the present study makes an original contribution to the field of SLA research.

1.2. Context of the study

As part of the data collection process, an interview with one of the Inspectors of English as a foreign language in Primary Education was conducted. As the interview data were judged to be appropriate for introducing the context of the study, all the information given in this section derive from that interview.

The study took place in primary schools in urban Nicosia, Cyprus. For the past fifteen years, the communicative language teaching (CLT) approach has been adopted for the teaching of English as a foreign language in primary schools. According to the inspector, the goals, set by the Ministry of Education and Culture in Cyprus, include developing the required skills for the children to be able to communicate effectively, freely and without any pressure in the target language. In addition, the lessons of English aim to cultivate positive attitudes towards acquiring the new language, as well as towards the culture associated with the target

language. In order to achieve these communicative goals, pair work and group work activities often feature in the current textbook.

At the time of the study, English was introduced as a subject in grade 4 (9-10 year-olds) and continued to be taught in grades 5 (10-11 year-olds) and 6 (11-12-year-olds). In all three grades, the subject was taught twice per week for forty minutes per lesson. The inspector referred to the time dedicated to English, as one of the weaknesses of English teaching in primary education. According to her, two 40-minute lessons per week are not enough to learn a language.

The majority of students in this context take private English lessons after school. Parents in Cyprus choose to offer these lessons to their children in order for them to be able to pass exams, which will serve as important qualifications in their future careers. The problem is that private language schools tend to adopt a more "traditional" structural approach to language teaching, as this is considered to be more effective with language exams. They usually teach grammar, spelling and ways to solve exercises. As the inspector argued in her interview, this approach contradicts the communicative approach followed in public schools. For her, this constitutes the second major weakness of language education in primary schools, as the effort of promoting communication in primary education is not cultivated by the whole language education received by these students.

In terms of the teachers of English in primary schools, these are often teachers who have a degree in primary school education and are expected to teach all subjects at all grades. In other words, the teachers are not specialised in foreign language teaching. As admitted by the inspector, there are not a lot of teachers, who are well qualified to teach English. Another problem is that most teachers do not want to teach English because they find teaching a foreign language a demanding task, when they do not have special training. Moreover, the

inspector also admitted that the Ministry has not trained the teachers well. The inspectors offer one seminar per year, which is attended by one teacher of English per school, rather than all teachers who teach the subject, due to practical reasons. The inspector refers to systematic training as the ideal, which however is not possible to be offered to these teachers.

Finally, it is important to mention the new curriculum that was prepared by the Ministry of Education and Culture and was implemented the academic year following the data collection of the present study, that is September 2011. This curriculum maintained the communicative goals of the previous curriculum, but differed in terms of the age these would be implemented. According to this curriculum, English would start being introduced as a foreign language from grade 1 in primary schools (6 to 7-year-olds). This introduction came only in oral and aural form through singing, story-telling and playing games. Students had these lessons once a week for forty minutes. As the inspector put it, these constituted "radical changes" to the existing curriculum.

As English was introduced in grade 1 for the first time, new material was required. While the inspector expressed her preference for packages of material that were already being published abroad by professional publishing houses, it was decided that the new material would be created and published in Cyprus, mainly for financial reasons. At the time of the interview, the materials for grades 1 and 2 had already been prepared by the Ministry of Education, while plans were made for the development of materials for higher levels. The books, that were being used at the time of the study in grades 4, 5 and 6, are going to be changed as well. According to the inspector, these books are dated and are not consistent with the current society in Cyprus, as they do not reflect its emerging multi-cultural aspect.

1.3. Overview of the thesis

There are a total of eight chapters in the present thesis. With this chapter briefly introducing the goals and the context of the study, Chapter 2 presents a review of the literature that revolves around four main research areas. These include task-based learning, classroom codeswitching with a particular focus on learner codeswitching, oral fluency and willingness to communicate. The discussion of empirical and theoretical studies in these four areas contributes to the formation of the hypothesis of the present study, which is presented at the end of this chapter.

Following the literature review, Chapter 3 describes the methodology of the study. This chapter includes the five research questions, the mixed-methods approach adopted, the description of the quasi-experimental research design, the sampling procedure, the ethical considerations and a brief description of the method of analysis for each research question.

The following three chapters concern the results of the present study. Chapter 4 presents the results of the qualitative analysis of teacher interview data and provides a response to the first research question. One of the main goals of this chapter is to introduce the context of the study in much more depth, with a particular focus on the role of codeswitching and task-based learning in this setting.

Chapter 5 presents the results of the qualitative analysis of student interactions. It aims to help the reader understand what was actually going on during the intervention, as well as to respond to the inquiries of the second research question.

Chapter 6 presents the results of the quantitative analysis of data coming from a pre-test, a post-test and a delayed post-test of the intervention. This chapter provides an answer to the final three research questions, which concern the different potential effects of the intervention.

The results of the three results chapters are summarised in Chapter 7, which is the Discussion chapter. In this chapter, the results are further discussed in relation to the results of previous relevant research. The chapter is divided into five main sections; each representing a research question.

Finally, Chapter 8 discusses the conclusions of the study, along with their implications for language teaching and learning. The chapter discusses the contribution of the study to certain areas of second language acquisition research, but by acknowledging its limitations. It ends by making recommendations for future research.

CHAPTER 2: Literature Review

2.1. Introduction

This chapter discusses theoretical and empirical studies conducted in four areas of research that are relevant to the present study. First, the discussion focuses on *task-based learning,* with particular reference to task repetition, as this plays an important part in the empirical part of this study. The second major section in this chapter concerns classroom codeswitching. While particular emphasis is placed on *learner codeswitching* studies, reference is made to more general issues, such as the debate on the use of the L1 for pedagogical purposes, studies on teacher codeswitching and on the perceptions of teachers' and students' on codeswitching. Then, the chapter reviews papers that concern *oral fluency*. Having defined what is perceived as oral fluency in the present study, the discussion focuses on how this develops with reference to theoretical frameworks, as well as to the different ways in which fluency has been measured in the past. Finally, the chapter refers to the concept of *willingness to communicate,* as an integral part of the present study. The discussion of these four areas of research leads to the hypotheses of the present study at the end of the chapter.

2.2. Task-based learning

As it will become evident later on from the review of studies on learner codeswitching and oral fluency, tasks constitute a very important aspect in the study of these areas, as most of them have been conducted within a task-based learning context. Learner codeswitching behaviour is usually analysed using task language, while fluency is often measured as participants complete tasks. Consequently, task-based learning plays an essential role in the present study. Task-based instruction is a type of teaching that is 'based entirely on tasks'

(Ellis, 2003: 351) and it derives from the Communicative Language Teaching (CLT) movement. CLT emerged as an opposition to the grammar-based approaches during the mid-1970s in the UK and in Europe and later on in North America (Samuda & Bygate, 2008).

As Richards (2002) argues, CLT was part of an effort to operationalise the notion of *communicative competence*. This notion emerged mainly from the sociolinguist's Dell Hymes' work on communicative competence (Hymes, 1974); an ethnographic work on communication which aimed to examine different perspectives of the participants' understanding of their interactive behaviour (Ellis, 2008). More specifically, Hymes (1974) argued that a child's ability to participate as a member in its society does not only involve speaking, but also communicating. Communicating refers to parameters such as the persons with which grammar should be used, the places that certain grammatical features are appropriate to be used, its purposes and so on. So when learning the system of grammar, the child also learns 'the system of its use' (Hymes, 1974: 75). In short, Hymes' notion of communicative competence entails that communicating is not just about correct language use; it is also about appropriate use of the language.

In an attempt to build on Hymes' (1974) theory, Canale and Swain (1980) proposed their own interpretation of communicative competence. Their definition of communicative competence included three types of competencies: *grammatical competence*, which involves demonstrating a mastery of basic grammatical principles during oral communication; *sociolinguistic competence*, which is knowledge about how language is used in different social contexts; and *strategic competence,* which involves employing communication strategies in cases of language difficulty or interactional breakdown (Canale & Swain, 1980; Shumin, 2002).

As mentioned above, communicative competence was operationalised as a language teaching approach with CLT. CLT gave rise to the belief that the ultimate goal of language learning should be successful communication and not formal accuracy (Cook, 2010). As Nunan (2004) argued, with CLT, language was viewed 'as a tool for communication' (7) and not as a set of 'phonological, grammatical and lexical items to be memorised' (7). Learners therefore started working with words and focusing on understanding the words of others (Cook, 2010). This movement replaced form-focused exercises with communicative activities, as well as teachers' authority with 'a new 'student-centred' curriculum' (Cook, 2010: 28). Teachers and students became *co-participants* in classrooms (Candlin, 1987), as learners were acknowledged as *active agents* (Samuda & Bygate, 2008).

During the rise of CLT however, uncertainty revolved around the place that explicit grammar should have in language instruction (Nunan, 2004). While some theorists believed that learners would be able to induce the grammatical knowledge of the language on their own through listening to comprehensible input (Krashen, 1982), others argued that focus on form instruction was necessary. During that time, Littlewood (1981) distinguished between the strong and the weak version of CLT. The strong version rejected focus on form instruction, and the weak version acknowledged its place in foreign language classrooms. Littlewood (1981) supported the weaker version of CLT because, as he argued, learners should possess linguistic competence and should be able to manipulate it based on the context of the interaction. More recently, researchers seem to agree with Littlewood's (1981) position, as there seems to be consensus that focus on form must have a place in the classroom (Nunan, 2004).

With this shift to the CLT approach, the types of activities also shifted from *analytic* to *holistic*. Analytic activities gave the opportunity to learners to concentrate on a single

linguistic structure and practise this structure to a greater extent (Samuda & Bygate, 2008). Examples would be drill exercises 'or exercises which involve the same operations from beginning to end' (Samuda & Bygate, 2008: 14). With CLT, holistic activities were favoured. These activities required learners to create meaning by using previous knowledge of phonology, vocabulary or syntax (Samuda & Bygate, 2008).

Tasks are one kind of holistic activities and their role in second language learning is to develop the learners' ability to use the new language (Samuda & Bygate, 2008). Samuda and Bygate (2008) place tasks very highly in language teaching methodology by arguing in their book that with holistic language work 'key language learning processes take place, and that tasks are invaluable in achieving this purpose' (8).

2.2.1. Defining tasks

Although there seems to be consensus on what the purpose of tasks should be, researchers have yet to agree on the definition of the term. Currently, various definitions of tasks exist in the literature. One of the earliest definitions was proposed by Prabhu (1987) who referred to the cognitive demands that tasks place on the learner. He argued that a task is 'an activity which require[s] learners to arrive at an outcome from given information through some process of thought [...]' (24). Another early definition came from Long (1985), who placed emphasis on the relationship of tasks with the world. Long (1985) defined tasks as 'things people will tell you they do if you ask them and they are not applied linguists' (89). Skehan (1996) defined tasks as 'activit[ies] in which: meaning is primary; there is some sort of relationship to the real world; task completion has some priority, and the assessment of task performance is in terms of task outcome' (38). Two years later Skehan (1998) added the presence of a communication problem to his existing definition. It is clear therefore that while

these definitions have some common references, they approach the concept of tasks from a different perspective; some emphasise the cognitive challenge of tasks, while others emphasise the importance of communicating meaning.

In an effort to interpret why so many different definitions of the term exist, Samuda and Bygate (2008) noted that definitions of tasks depend on the perspectives of the person who is defining tasks and on the purposes a definition will serve. Taking into account this variety in the literature, Samuda and Bygate (2008) chose to create their own definition by considering the features that a task *must* have, rather than those that a task *could* have. They came up with the following working definition:

> *'A task is a holistic activity which engages language use in order to achieve some non-linguistic outcome while meeting a linguistic challenge, with the overall aim of promoting language learning, through process or product or both'* (Samuda & Bygate, 2008: 69).

Their definition therefore suggests that, when completing a task, learners should try to accomplish a non-linguistic outcome and should be linguistically challenged in their effort to do so. In other words, their focus should be on achieving the outcome of the task and through this process language is expected to be developed.

Rather than offering a definition, some researchers preferred to offer criteria or key features that tasks should have. The first criterion which seems to be crucial for all researchers is *focus on meaning* (Skehan, 1998; Ellis, 2008, 2009). This suggests that learners should concentrate more on the pragmatic and semantic meaning of utterances, rather than on how grammatically correct they are. Secondly, there should be some kind of *'gap' of information*, which should be filled in by the learners through an exchange of information (Ellis, 2008, 2009). Third, there should be a clearly defined outcome as the goal of the task and it should be different to the use of language (Ellis, 2008, 2009). As it is suggested by Samuda and Bygate's (2008)

definition, language should serve as the means to accomplish this outcome and not the end (also in Ellis, 2009). Lastly, the language produced in tasks should be authentic, in the sense that it should resemble the kind of language that could be encountered outside the classroom (Skehan, 1998; Ellis, 2003). Finally, learners should rely on their own linguistic, as well as non-linguistic, resources to complete the task (Ellis, 2008, 2009).

It is also worth noting that Ellis (2009) distinguishes between two types of tasks, *focused* and *unfocused tasks.* Focused tasks provide opportunities to communicate using specific linguistic features, such as a grammatical structure (Ellis, 2009). Learners might not be explicitly told which grammatical structure they will be practising with that task. In unfocused tasks however, learners use the 'language in general communicatively' (Ellis, 2009: 223).

From the discussion above, it seems that while the lack of a clear definition of tasks is still an issue, researchers seem to agree on the role that tasks play in language teaching, as well as on the features and purposes tasks should have. Tasks are essential for language communication which is authentic and meaningful. When completing tasks, students should aim to achieve a certain non-linguistic goal, by challenging themselves in the target language.

2.2.2. Teachers' opinions on tasks

Apart from the opinion of researchers on tasks, it is essential to look at the opinions of practitioners on tasks. Our knowledge in this area comes mostly from the work of Carless and colleagues, who extensively examined the opinions of teachers on task implementation in Hong Kong (Carless & Gordon, 1997; Carless, 2002, 2003, 2004, 2008). In his earliest study (Carless & Gordon, 1997), the opinions of 89 primary school teachers on task-based teaching were examined before and after they implemented tasks in their classrooms for the first time.

These teachers completed a questionnaire at these two time points, while ten of them were randomly chosen to participate in interviews. The main finding of the study was that the teachers faced much fewer difficulties with task implementation than they initially anticipated. This finding was confirmed in the survey conducted by Clark *et al.* (1999) in the same context, who found that, overall, teachers had positive attitudes towards task-based teaching and that their positive attitudes were enhanced after they implemented this new approach.

The difficulties teachers reported facing in Carless and Gordon's study (1997) included the pupils' use of Cantonese, which the teachers regarded as their own personal 'failure'; lack of time, which was required both in order to prepare tasks, as well as to execute them in class; large classroom size; and instructions and explanations, which had to be given in the L1 in order for students to understand. Despite these difficulties, only 6% of the teachers reported having negative feelings towards this new teaching approach, while 96% thought that their students would benefit from it. Although they did not reject task-based learning, the majority of the teachers reported that they did not believe that the benefit gained from tasks would be greater than the benefit gained from other types of activities. In other words, the task-based approach provided equal learning opportunities as any other activities. In terms of the students, all teachers reported that the students enjoyed the experience and found what they were doing very interesting.

In a later study, Carless (2002, 2003) observed and interviewed repeatedly three primary school teachers in Hong-Kong over the course of an academic year. The teachers were also asked to complete an attitude scale, which formed the primary source of data on teachers' attitudes towards task-based learning. In general, the teachers had positive attitudes towards task-based teaching because students, not only enjoyed tasks, but they got involved with the

language by working in groups. Difficulties they reported facing included lack of discipline (i.e. making noise by talking among them) when students engaged in pair work. The large classroom size also contributed to the lack of discipline, as group work caused a lot of noise.

In addition, teachers in Carless' study (2002) reported that the students' use of the L1 was one of the main problems when carrying out tasks. Students tended to use Cantonese very frequently when in groups, either to discuss something relevant to the tasks, but also something irrelevant. Carless (2002) argued therefore that '[w]hile this was a normal social function, it was unlikely to promote English language learning' (392). Lack of time was another difficulty found in the teachers' reports. Teachers felt the pressure for completing the syllabus on time, and tasks were time-consuming both in terms of their preparation, as well as their implementation. This was also a finding reported by Li (1998) who examined the perceptions of 18 secondary school teachers in the South Korean context. Here, students were aimed towards passing exams and time therefore was mostly devoted to developing test-taking skills. Similarly, in the context of Carless' study, the policy of the schools urged teachers to spend more time on reading, writing and grammar, rather than on oral activities. It could be argued therefore that the barriers faced by the teachers are more of a practical nature rather than a pedagogical nature; teachers' issues with tasks were not related to the value of tasks, but to their in-class implementation.

A few years later, Carless (2008) looked into the issue of L1 use during task completion in more depth by interviewing ten secondary school teachers in Hong Kong. The majority of the teachers reported that their students' L1 use was inevitable. Although this was not preferable, they acknowledged that it helped students maintain their attention, interest and involvement. Teachers reported students' proficiency level to be a key issue. More proficient students would be able to maintain TL use much more than lower proficiency students. Lastly, the

teachers referred to several strategies that would help students increase their TL output, including assigning the role of 'language monitors' to some students in order to remind their peers to use the TL; further assigning these students to write down the L1 utterances and then having these translated into the L2 as a follow up activity; use whole-class tasks, rather than pair or group work tasks; or use task repetition. From the data, Carless (2008) concluded that L1 use seemed to be a 'humanistic and learner-centred strategy, with potential to support student learning' (336). He warned however that it could also prevent TL practice and communication.

More recently, Andon and Eckerth (2009) also explored teachers' attitudes towards task-based teaching. Four student teachers, studying for a master's degree in ELT, were involved in multiple interviews and observations, as well as in a stimulated recall process based on the researcher's description of their lessons. From the interviews, it became evident that the teachers used tasks for four purposes. First, they used them for communication, as they provided students the freedom to create their own language. Second, tasks were used 'as contexts of situational and interactional authenticity in L2 use' (Andon & Eckerth, 2009: 293). This means that teachers used them as opportunities to establish connections between the language used in the classroom and the authentic language used outside the classroom. Third, tasks were viewed as activities which set goals and had an outcome. Teachers' viewed this as an advantage, as their students focused on achieving the goal of the task, rather than on using the language itself. Language therefore came 'naturally'. Finally, tasks were used 'as reference points for form-focused activities' (Andon & Eckerth, 2009: 300). In other words, teachers used them in conjunction with other more form-focused activities. As Andon and Eckerth (2009) put it, this could be characterised as task-*supported* teaching, rather than as task-*based* teaching. In conclusion, although teachers believed that tasks could help students

notice language and use their own language creatively, they thought that tasks could not act as 'the main driver for the acquisition of new language' (Andon & Eckerth, 2009: 304).

In conclusion, although teachers' opinions on task implementation were not studied widely, the evidence from studies in different contexts seems to be similar. While teachers can see the importance of using tasks in their classrooms due to its social and cognitive benefits, they often refrain from using them extensively as their main teaching approach. This is due to both practical, as well as pedagogical reasons. Practical reasons would include lack of time and difficulty of maintaining good discipline. Pedagogical reasons would include the teachers' concerns regarding their students' excessive L1 use. Task implementation therefore does not come without problems in these contexts.

2.2.3. Empirical studies on task-based learning

Research on task-based learning has largely focused on how the three components of oral production (accuracy, complexity and fluency) were affected as task outcomes by the manipulation of a range of variables. These variables may concern the structure of the task (i.e. planning time, single or multiple goals) or they could be learner-related (i.e. topic familiarity, interlocutor familiarity).

Robinson's (2005) *Cognition Hypothesis* is largely related to this area of research, as it suggested that tasks should be sequenced in increasing cognitive complexity, in order for more complex and grammaticised output to be produced. Having categorised these variables, Robinson (2005) created the *Triadic Componential Framework,* shown in Figure 2.2.1 below. First, the *task complexity* category includes variables that are relevant to the cognitive demands of a task, such as planning time or topic familiarity. Second, the *task condition*

category includes interactive factors involved in pair work, such as interlocutor familiarity and shared proficiency. Third, the *task difficulty* category includes variables relevant to learner abilities, such as anxiety, working memory capacity and language aptitude. In designing tasks, therefore, one should take into consideration these variables, depending on the proficiency of the learners and the aims of the task.

Task Complexity (Cognitive factors)	***Task Condition*** (Interactive factors)	***Task Difficulty*** (Learner factors)
(Classification criteria: cognitive demands)	(Classification criteria: interactional demands)	(Classification criteria: ability requirements)
(Classification procedure: information-theoretic analyses)	(Classification procedure: behavior-descriptive analyses)	(Classification procedure: ability assessment analyses)
(a) **Resource-directing variables** making cognitive/conceptual demands	(a) **Participation variables** making interactional demands	(a) **Ability variables** and task-relevant resource differentials
+/– Here and now	+/– Open solution	h/l Working memory
+/– Few elements	+/– One-way flow	h/l Reasoning
–/+ Spatial reasoning	+/– Convergent solution	h/l Task-switching
–/+ Causal reasoning	+/– Few participants	h/l Aptitude
–/+ Intentional reasoning	+/– Few contributions needed	h/l Field independence
–/+ Perspective-taking	+/– Negotiation not needed	h/l Mind/intention-reading
(b) **Resource-dispersing variables** making performative/procedural demands	(b) **Participant variables** making interactant demands	(b) **Affective variables** and task-relevant state-trait differentials
+/– Planning time	+/– Same proficiency	h/l Openness to experience
+/– Single task	+/– Same gender	h/l Control of emotion
+/– Task structure	+/– Familiar	h/l Task motivation
+/– Few steps	+/– Shared content knowledge	h/l Processing anxiety
+/– Independency of steps	+/– Equal status and role	h/l Willingness to communicate
+/– Prior knowledge	+/– Shared cultural knowledge	h/l Self-efficacy

Figure 2.2.1 The Triadic Componential Framework (Robinson, Cadierno & Shirai, 2009: 535)

One of the variables that have been largely investigated in the past is *planning time* (e.g. Foster, 1996; Foster & Skehan, 1996; Skehan & Foster, 1997). These studies examined whether having a pre-task phase, during which learners are given time to prepare the solution

of the task, affects the learners' performance. In a very early study with twenty Japanese EFL learners, Crookes (1989) found that planning time had a positive effect on complexity and fluency, but not on accuracy. Similarly, Foster and Skehan (1996) reported that in their study unguided planning, that is solitary planning time, led to greater fluency and complexity. Also, guided planning led to even greater complexity and fluency than unguided planning. As for accuracy, Foster and Skehan (1996) reported that unguided planning resulted in the greatest accuracy gains. The authors' interpretation was that unguided planners used this time to practise language, while guided planners made the task even more complex than it was. Evidence of the positive effect of planning time on task performance was provided by other studies as well (e.g. Wigglesworth, 1997).

There is also limited empirical evidence suggesting that *interlocutor familiarity* affects task performance. In a study with 32 adult Japanese learners of English, O'Sullivan (2002) reported that when students worked with a familiar interlocutor, they achieved higher scores (given by raters), than those achieved while working with an unfamiliar person. Hancock's (1997) study, which will be discussed in greater detail later on in the learner codeswitching section of this review, suggested that choosing interlocutors may act as an incentive for students to get involved in the task more than they actually would. In the same study, *topic familiarity* emerged as another variable affecting students' performance. It is important therefore to consider such variables when designing and preparing tasks, as research evidence suggests that these affect learners' performance.

Another study conducted by Iwashita (2001) examined the effect of proficiency level in learner interactions during tasks. The participants, who were 24 university students in Australia learning Japanese, were allocated into three different groups: a low-low proficiency group (LL), a high-high proficiency group (HH) and a low-high proficiency group (LH).

These pairs were audio-recorded while completing three different tasks: a jigsaw task and two information-gap tasks. In order to determine the effectiveness of the interactions, the amount of opportunities for modified output (i.e. clarification requests and confirmation checks) were counted. Iwashita (2001) found more interactional moves in LH proficiency pairs than same-proficiency pairs, but this did not lead to more modified output. Nevertheless, this study suggests that mixed-proficiency pairs can have more effective interactions than same-proficiency pairs.

2.2.4. Task repetition

Another aspect of task-based learning research, which is of particular importance to the present study, is task repetition. What is meant by task repetition is the 'repetition of the same or slightly altered task – whether the whole tasks, or parts of a task' (Bygate & Samuda, 2005: 43). According to Bygate (2001), who seems to be the earliest and most prominent supporter of task repetition, when speakers first complete a task on a familiar topic, their main concern is 'to find appropriate formulations, and articulate them' (Bygate, 2001: 26). At this stage, speakers are more concerned with conducting a meaningful message that they would like to convey, rather than formulating this perfectly with precise vocabulary and correct grammar (Bygate, 1996). On the second occasion however, speakers draw on the concept and encoding of information they had previously used. Their attentional resources therefore, are freer to focus on other aspects of oral production, such as accuracy, complexity and fluency. Bygate (2001) noted that the minimum that could happen with task repetition is an increase of fluency.

One of the earliest studies investigating task repetition was a pilot case study conducted by Bygate (1996). In this study, a learner performed a task, which involved watching a silent

cartoon video and then narrate the story. Three days later, the learner repeated the same task without being aware that this repetition would take place. Bygate (1996) found that the fluency and accuracy (vocabulary, idiomaticity, grammatical markers, structure) of the learner were improved massively. In terms of fluency, Bygate (1996) reported that she used repetitions in order to find suitable formulations, while on the second occasion, she used repetitions in order to check formulations. In terms of complexity, the learner had a 75% increase of subordinate clauses. Bygate (1996) concluded therefore that these results supported the hypothesis that on the first occasion the learner was concerned more with content, while on the second occasion the focus was on the linguistic formulation because the content was more familiar to her. Although the study provides an indication of the potential of task repetition, the results cannot be generalised as there was only one participant.

Another study on task repetition was conducted by Gass *et al.* (1999), who examined the effects of task repetition on the linguistic output of 103 English-speaking learners of Spanish at a university in the US. The participants were randomly assigned into three groups: two experimental and a control group. During the intervention, all groups were asked to watch silent films with the character of Mr. Bean and then narrate the story while being recorded. The first experimental group watched the same video three times, while the second experimental group watched three different videos each time; both groups with a two or three-day interval. On time 4, which was a week after time 3, both groups watched a new video. The control group only watched the time 1 and time 4 videos. Figure 2.2.2 below presents the design of the study.

Group	Time 1	Time 2 2–3 days later	Time 3 2–3 days later	Time 4 1 week later
Same Content ($n = 32$)	Library	Library	Library	Packing
Different Content ($n = 33$)	Library	The guard	Lunch	Packing
Control ($n = 38$)	Library			Packing

Figure 2.2.2. Gass *et al.* (1999) intervention design (p. 556)

The data were analysed in terms of overall proficiency, morphosyntax and lexical sophistication (advanced words that are not on the list with the 200 most commonly used Spanish words). Gass *et al.* (1999) found some evidence supporting the improvement of the first experimental group who watched the same video repeatedly. More specifically, the same-content group showed greater accuracy and they increased their lexical sophistication by using less common words more often. No evidence supported however, the transfer of these effects on a new task.

The problem with this study is that the control group received no treatment during the intervention, while the experimental groups had more exposure to tasks and more opportunities to practise the language. Comparing their performance therefore becomes somewhat unequal, as the control group was at a disadvantaged position. Perhaps if this group received some kind of exposure to the L2, without necessarily practising tasks, then their performance would have been more comparable.

A study that used an interesting methodology for investigating task repetition was conducted by Lynch and MacLean (2000, 2001). Lynch and MacLean asked their participants to complete a task they designed, called the *poster carousel*. The participants, who were students taking courses of English for Academic Purposes, were asked to read an academic medical article and prepare a poster presentation on it. Each student then had six different ‘visitors’ to

their poster. As each visitor asked their own questions, the task of presenting the poster could not have been exactly the same each time; it was 'more like *recycling,* or *retrial'* (Lynch & MacLean, 2000: 227). Having recorded all the conversations between poster-presenters and visitors, Lynch and MacLean found that this repetition had positive effects on the participants' accuracy, which was measured by counting the amount of correct and incorrect structures in terms of syntax (subject-verb structure), lexico-grammar and phonology (pronunciation).

However, one of the main differences of this study to the others that investigate task repetition is the input material. In this study, the participants received input from an 800 to 1000-word medical article, while participants in other studies received input that contained no linguistic cues. Moreover, in Lynch and MacLean's (2000) study, the participants were given long planning time (60-70 minutes) to prepare their presentations, while in the other studies the participants were hardly given any planning time. The positive effect of task repetition in this study therefore should be considered in relation to these variables, as it is hard to compare the results of this study to those of other similar studies in the field.

Following up on his previous studies on task repetition, Bygate (2001) compared the performance of 48 non-native speakers of English university students on two sets of tasks: a narrative set and an interview set, each consisting of six different versions. As in his previous study, the narrative tasks involved watching silent cartoon videos and then narrating the story while being recorded. The interview tasks were structured around pictures which reflected some aspect of life in Britain. The participants were randomly allocated into three groups: the narrative group, the interview group and the control group. The 10-week intervention consisted of five time points with a two-week interval. At time 1, all three groups completed the same narrative and interview task. On weeks 3, 5 and 7, the experimental groups

completed two tasks on each session: a repetition of the previous session's task and a new task. During this period, the control group received no treatment. Finally, on week 10 all three groups completed the same narrative and interview tasks they completed on week 1, as well as a new task for each type of task. The data of the study came from the performances recorded on week 10 only.

Group	Treatment 1	Treatment 2			Data
	Time 1 Week 1	Week 3	Week 5	Week 7	Time 5 Week 10
Narrative group	Nar. 1 Int. 1	Nar. 2 & 3	Nar. 3 & 4	Nar. 4 & 5	Nar. 1 & 6 Int. 1 & 6
Interview group	Nar. 1 Int. 1	Int. 2 & 3	Int. 3 & 4	Int. 4 & 5	Nar. 1 & 6 Int. 1 & 6
Control group	Nar. 1 Int. 1				Nar. 1 & 6 Int. 1 & 6

Figure 2.2.3. Bygate's (2001) intervention design (p.31)

Investigating the effect of *task-type practice* first, Bygate (2001) compared the performance of the experimental groups, who were exposed to tasks, to that of the comparison group, who had no exposure to tasks. The performances were compared in terms of fluency (number of filled and unfilled pauses per t-unit), complexity (number of words per t-unit) and accuracy (number of errors per t-unit). While there was no clear evidence that practising a certain type of task was beneficial, Bygate found that being exposed to the same task for a long period of time, rather than just once, had a significant positive effect on fluency. Bygate concluded however that task repetition does not contribute to acquisition, since the practice effect did not seem to be transferable.

Secondly, Bygate examined the effect of *task repetition* by comparing the performance of the task that was performed in week 1 and then repeated in week 10, to the performance on a new version of that task (week 10). He found a significant effect on fluency and complexity, but

not on accuracy. Finally, Bygate looked at the *task effect,* or else the effect of the type of task on oral performance. He found that the output of all groups was significantly less fluent, but more complex, on the interview tasks in comparison to the narrative tasks. One explanation given by Bygate was that the interview tasks required more pausing than the narrative tasks due to its interpersonal dimensions.

However, this study did not come without problems. The measures used for fluency, accuracy and complexity may not be representative enough. These are complicated constructs which had been measured with a combination of measures in the past. Simply having one measure for each construct might not capture the full range of aspects that fluency, accuracy and complexity represent. Fluency, for example, has been measured in terms of speech rate, mean length of runs, pause measurements, as well as dysfluency measurements in the past. Counting the number of filled and unfilled pauses per t-unit alone therefore might not be a valid fluency measure. Moreover, another methodological problem concerns the unequal comparison between the experimental and the control group, as in the Gass *et al.* (1999) study. The experimental groups were bound to do better than the control group, as they received much more exposure and much more opportunities to practise the L2, than the control group, who received no treatment of any kind during the 10-week period. Perhaps if the control group received some kind of exposure to the language during this period, not necessarily by completing tasks, then the progress of the experimental and control groups would have been more comparable.

In a more recent study, Ahmadian (2011) followed Bygate's (2001) investigation on the effect of *task-type practice,* by examining whether the effect of 'massed repetition' of the same task carries over to the performance of a new task. According to Ahmadian (2011), 'massed repetition' is the repetition of the same task for several times over a long period of time. In his

intervention study, Ahmadian (2011) used two intact classes, each consisting of fifteen female learners of English in a language centre in Iran. Having ensured the homogeneity of the groups with the Oxford Placement test, one class acted as an experimental group and the other as a control group. In week 1, both groups were given a *dialogic narrative task* to complete which, as in Bygate's studies, involved watching a silent film and narrating the story to a researcher without time for planning. After this initial task, the experimental group was given the same dialogic narrative task to complete on ten more occasions, with a two-week interval between each occasion. During this time, the control group received no treatment. Then on time 12, both groups completed an interview task, during which they were asked questions about their personal experience of language learning, as well as their opinion on the importance of knowing a second language. The intervention lasted for six months. The participants were audio-recorded while completing the tasks. The table that follows presents graphically my understanding of the stages of their intervention study.

Table 2.2.1 Graphic representation of Ahmadian's (2011) intervention design

	Occasions (6 months) 2-week intervals between each occasion							
	1	**2**	**3**	**4**	**...**	**10**	**11**	**12**
Experimental	Dialogic narrative task	Dialogic narrative task	Dialogic narrative task	Dialogic narrative task	...	Dialogic narrative task	Dialogic narrative task	Interview
Control	Dialogic narrative task	/	/	/	...	/	/	Interview

Ahmadian (2011) compared the performance of the two groups on time 1 and time 12 based on the three components of oral production. Complexity was measured with the amount of subordination, verb forms and the mean length of AS units. Accuracy was measured with the percentage of error-free clauses and the percentage of correct verb forms. Fluency was measured with syllables per minute and syllables per minute excluding dysfluency markers.

At time 1, the researchers found no significant differences between the two groups on any of the measures. At time 12 however, the experimental group did significantly better on all measures of fluency and complexity.

Ahmadian (2011) concluded therefore that massed repetition can contribute to the development of interlanguage. This claim however, seems to be somewhat too ambitious. As in the studies of Gass *et al.* (1999) and Bygate (2001), the problem here is that the control group received no kind of treatment, while the experimental group had more opportunities to practise the L2 during this long period of time. If the control group attended extra sessions where they also practised their speaking without completing tasks, then it would have been a more equal comparison.

Ahmadian and Tavakoli (2011) examined the effect of online planning and task repetition on the oral production of 60 female adult learners of intermediate-level English in a language centre in Iran. These students were randomly allocated to four conditions: careful online planning with task repetition, careful online planning without task repetition, pressured online planning with task repetition and pressured online planning without task repetition. The learners under careful online planning conditions were given as much time as they needed in order to prepare the task, while the learners under pressured online planning conditions were given a maximum of six minutes for planning. Examining the effects of these conditions on accuracy, complexity and fluency, the authors found that the condition of careful online planning with task repetition significantly enhanced all three components of oral production. This was not a surprising result, however, as this was the group that received the most additional help, i.e. they repeated the task and they were allocated planning time to prepare for it.

The most recent study on task repetition differs from the others described above, in the sense that it tests a task repetition plus feedback package, rather than task repetition alone. This study was conducted by Baleghizadeh and Derakhshesh (2012) and it aimed to examine the effects of task repetition with feedback on the learners' output. Four Iranian young adult female learners of English (intermediate level) were audio-recorded while giving a fifteen to twenty-minute oral presentation on a book they had read. The participants were then asked to transcribe their presentations using the audio recordings, correct their own mistakes and submit this to their teacher for additional corrections. These corrections focused on verbs, preposition usage and direct translation only. Having the revised draft returned to them, the participants were asked to give a second oral presentation.

Comparing the erroneous utterances of the two presentations, Baleghizadeh and Derakhshesh (2012) found that the second oral performance was more accurate and concluded that task repetition plus feedback can have a positive effect on accuracy. However, the possibility of the students memorising the correct utterances from the revised draft for the second presentation cannot be disregarded. This study is closely related to the present study due to the task repetition plus feedback package it incorporates.

Concluding the section on task repetition, it seems that studies in this area are limited and very similar. Under laboratory conditions, experimental groups narrate the story of a silent film repeatedly, while control groups do not receive any opportunities for such practice. This constitutes the main methodological problem of these studies, because it makes the groups non-comparable. In addition, sample size in these studies tends to be relatively small and this limits generalisation prospects. Despite these problems, there is some evidence suggesting that task repetition has a positive effect on fluency and complexity (e.g. Bygate, 2001; Ahmadian, 2011). The measures, with which these have been examined however, could be

criticised as being too simplistic and unable to capture therefore the complex nature of these constructs.

What researchers in this area aim to find is that the benefits from same-task repetition can be transferred to the performance of a new task (Ahmadian, 2011). This kind of evidence would be very significant, as it would suggest that task repetition positively contributes to language learning. However, no empirical study so far has been able to provide such evidence. In addition, considering the pedagogical aspect of task repetition, it seems somewhat hard to imagine the implementation of pure task repetition in real classroom contexts. With the context of the present study in mind, task repetition is not always viewed positively by teachers or young learners, as the qualitative analysis will show later on. Young learners wish to work on new material, unless they are given a good reason about why they are repeating the same activities. What I would like to propose therefore, is the idea of task repetition *plus* feedback, rather than task repetition alone. This package may have two kinds of effects: first, the feedback component could be used as a valid justification of why a task should be repeated; and second, teachers might find repeating tasks with incorporated feedback more pedagogically acceptable. As in Baleghizadeh and Derakhshesh's study (2012) therefore, the present study uses the task repetition plus feedback package, and its benefits are compared to no task repetition.

2.2.5. Conclusions on task-based learning

Task-based learning has become a very influential approach in language teaching methodology. However, the discussion so far has shown that it does not come without problems. These problems stem from both research, which has yet to agree on a definite term of what a task really is, as well as from practical issues with task implementation reported by

teachers (e.g. lack of time, L1 use, lack of discipline). In terms of task repetition, the limited number of studies in this area provided some evidence of a positive effect on fluency and complexity. The methodological problems of these studies however, should not be disregarded. Finally, a task repetition *plus* feedback package is proposed as a more practically acceptable approach in the present study.

2.3. Codeswitching

The second major research area reviewed in the present chapter is that of classroom codeswitching. Having initially defined the term, the reasons why this has been an issue of controversy are discussed, while the focus of empirical studies in this area follows.

2.3.1. Defining codeswitching

Codeswitching has been defined as the ability to use two languages alternatively in discourse (Poplack, 1980). While other terms were used in the past to describe the same phenomenon, such as “codemixing”, “codeshifting” and “language switching” (Benson, 2001), the term codeswitching is the one that is most widely used and accepted. According to Benson (2001), this term was initially used sometime in the 1950s, although its exact origin is still not clear.

Codeswitching has been studied widely as a phenomenon in two different fields: bilingual studies and SLA studies (Arnfast & Jorgensen, 2003). In the field of bilingual studies, codeswitching is studied in naturalistic contexts and it is viewed as an asset and a valuable addition of bilingual speakers (Macaro, 2005). In a very early study on naturalistic codeswitching, Poplack (1980), who followed a bilingual Puerto Rican community in New York City, identified three different types of codeswitching. The first was that of *inter-sentential switching,* occurring between clauses. More specifically, inter-sentential switching

is when a whole sentence is expressed in one language and the following sentence occurs in another language. The second type of codeswitching was *intra-sentential codeswitching,* occurring within clauses. An intra-sentential switch is when a sentence begins in one language and finishes in another. Finally, the third type of codeswitching identified by Poplack (1980) was *tag-switching,* which refers to instances where a tag of a different language is included. This is similar to inter-sentential codeswitching, with the difference that the tag is free of syntactical constraints. However, the problem with this distinction of types of codeswitching is that they refer to clauses or sentences, rather than utterances that people usually speak in. In naturalistic contexts, it is less likely to encounter conversations that consist of full sentences.

Research on naturalistic codeswitching has also attempted to investigate whether this phenomenon occurs within some kind of grammatical boundaries. The most widely known model is the *Matrix Language-Frame Model,* which accounts for intrasentential structures only (Myers-Scotton, 1997). This model suggests that one of the languages of switching is the *matrix language*, or else the dominant language that 'sets the morphosyntactic frame' (Myers-Scotton, 1997: 3) of codeswitching sentences. The other language is the *embedded language*, which has a smaller role in the codeswitching behaviour (Myers-Scotton, 1997).

As the present study concerns codeswitching occurring within FL classrooms, this literature review focuses on studies from SLA research. Nevertheless, it is worth mentioning here that it has been argued in the past that the gap between naturalistic and classroom codeswitching should be re-examined and re-evaluated. In a study with learners of Danish, Arnfast & Jorgensen (2003) found evidence suggesting that codeswitching can be used both as a communication strategy, as well as a competence or resource that contributes to acquiring a language and being socially accepted. Their data further suggested that the role of codeswitching depended on proficiency level. While at beginner stages learners used

codeswitching as a compensation strategy, as they progressed, their codeswitching behaviour was developing into a bilingual competence. This competence was not used due to lack of proficiency, but it was used consciously in order to signal something; in other words, it was being used in the same way bilinguals use it. Having established this link between the two types of codeswitching, the discussion will now focus on classroom codeswitching.

2.3.2. Issue of controversy

Classroom codeswitching had become an issue of controversy for many years in SLA research, as many researchers as well as practitioners did not always see its use positively. It was often believed that learners' L1 use is a sign of poor L2 proficiency. Similarly, FL teachers reported having feelings of guilt when they resorted to L1 use, as they viewed it as *recourse to L1* (Macaro, 2005).

Looking at the history of foreign language teaching, the roots of this negative connotation attached to L1 use can be found in the *Berlitz method.* This was the method used in the popular *Berlitz schools* in the early 1900s, which were private commercial language classrooms for adult language learners (Cook, 2010). The schools were initially established in the USA in 1877 (Berlitz, 2012) for immigrants who wished to learn English fast in order to facilitate their new lives there, and then expanded in Europe for traders and tourists, who wished to learn European languages (Cook, 2010). Within the Berlitz schools, who took their name after their founder Maximilian Berlitz, the so-called Berlitz method was followed. This method was one of the first rejection of L1 use in language teaching methodology. It focused on speaking and promoted teachers who were native-speakers of the target language (Cook, 2010).

Around the same time the Berlitz method was flourishing, that is in the 1990s, the *Reform Movement* emerged. This was an effort of language teachers and linguists to signal the need for new language teaching approaches (Richards & Rodgers, 2001). This movement differed to the Berlitz method in the sense that it was developed for academic and pedagogic purposes in secondary schools, while the former concerned purely commercial purposes (Cook, 2010). The adult clients of the Berlitz schools were paying and therefore demanding fast results, whereas this pressure did not exist in secondary schools.

Although these two streams developed separately and for different purposes, they merged and created 'a strong and coherent new programme for language teaching' (Cook, 2010: 7), namely the so-called *Direct method.* Although it is still not clear who created the term, Cook's (2010) definition of the Direct method involves 'all teaching which excludes use of the students' own language from the classroom, whether for translation or for explanation and commentary' (7). This movement away from the students' native language towards teaching that involved doing everything in the target language constituted the first revolution in the history of language teaching methodology (Cook, 2010).

The second revolution in language teaching methodology occurred in the 1970s with the rise of two new streams: the *Natural Approach* and *Communicative Language Teaching (CLT).* The Natural approach, which was the source of early SLA theories, viewed learning as 'an automated internal biological process' (Cook, 2010: 27), which simply needed what Krashen (1982) referred to as 'comprehensible input' in order to subconsciously activate the cognitive processes required for language acquisition (Cook, 2010). According to this approach, the use of the learners' L1 was not necessary in language teaching, as meaning could be conveyed through demonstration (Richards & Rodgers, 2001). Learners would be able to induce grammatical rules through spontaneous talk in the classroom (Richards & Rodgers, 2001). Similarly, as it was discussed in the task-based learning section, CLT put forward the idea

that the ultimate aim of language learning should be successful communication. The acquisition of forms therefore would only take place in order to fulfil the need for meaningful communication (Cook, 2010). Both the Natural Approach in SLA research and CLT implied a primary role for communication and conveying meaning, and a secondary place for focus on forms.

During the period of the second revolution in language teaching, the use of the learners' L1 in the classroom was not so much rejected, but ignored (Cook, 2001). The largely influential input theories in SLA research implied L2-only instruction. As already mentioned, Krashen's (1982) controversial *Comprehensible Input hypothesis* placed *comprehensible input,* that is language which contains structures a bit more advanced than the learners' current level of competence ($i + 1$), as a primary condition for L2 acquisition. Krashen's theory was heavily criticised however mainly for being unclear and for lacking empirical evidence (Gregg, 1984). For example, Gregg (1984) argued that Krashen used the $i + 1$ in an unclear way; this sometimes referred to 'the next structure to be acquired' (87) and other times it referred to 'the learner's competence at one 'stage' after *i*' (87).

Following his claim on $i + 1$, Krashen further argued that formal L2 instruction should just serve as a source of comprehensible input without necessarily focusing on form. This position was criticised mainly by McLaughlin (1987) who argued that it ignored 'the advanced cognitive development of adults and the advantages of formal teaching and learning' (46). Adult learners have certain cognitive skills which enable them to use formal instruction to their advantage. Krashen's view on formal instructions would deprive them of using these skills. Furthermore, McLaughlin (1987) also pointed out the lack of empirical evidence in Krashen's theory. Krashen validated his theory by describing how his theory explained certain phenomena (e.g. silent period). According to McLaughlin (1987) however, this strategy

cannot be considered as real evidence for theory validation. Finally, other theories on input and interaction followed that of Krashen's. Long (1985) stressed the importance of learner involvement in the interaction by negotiating meaning, while Swain (1985) added language production as an essential condition for acquisition.

Despite these developments, the value of TL input was never doubted. These influential theories implied L2-only instruction and consequently did not allow any space for the L1 in the classroom. As a result, a debate among researchers regarding the value of TL exclusivity began. Some researchers argued that learners do not need to understand everything that is being said in the classroom (e.g. Chambers, 1991; Halliwell & Jones, 1991). Trying to 'figure out' what the unknown parts mean constitutes the basis of language learning (Wong-Fillmore, 1985). Others stressed the importance of TL input by indicating that any use of the L1 cuts down on the amount of valuable TL input (Ellis, 1984). As this debate featured at the time of the rise of CLT, some researchers argued that L1 use was inappropriate, as it 'runs counter to the promotion of authenticity' (Atkinson, 1993: 2).

On the other side of the debate, there were theorists who argued that no empirical evidence showed a negative effect of the L1 use on TL proficiency. These commentators regarded the L1 as a useful pedagogical tool that could be used as a resource (Cook, 2001). Butzkamm (2003) adopted a much stronger position in favour of the mother tongue when he argued that it 'is *the* master key to foreign languages, the tool which gives us the fastest, surest, most precise, and most complete means of accessing a foreign language' (31). He further argued that the L1 is the learner's "strongest ally" that should be used systematically (Butzkamm, 2003). In a study on learner codeswitching, Swain and Lapkin (2000) characterised the L1 as an 'important cognitive tool' (269). In terms of the pedagogical aspect, Atkinson (1987)

argued that ignoring the L1 'is almost certainly to teach with less than maximum efficiency' (247).

More recently, researchers seem to acknowledge that the L1 can be used as a valuable pedagogical tool that facilitates language learning (Macaro, 2001; Cook, 2001), as no recent publication supports TL exclusivity. Researchers however called for the construction of a framework that will indicate to practitioners when codeswitching is beneficial and when it is simply the easy option (Macaro, 2001). This framework will ensure the 'judicious' use of codeswitching which will have a facilitative and valuable role in the classroom, rather than a hindering one.

Despite the fact that researchers seem to have reached a consensus regarding the issue of codeswitching, textbooks and policymakers around the world still seem to promote TL exclusivity. A recent example from the Korean context comes from Lee (2010), who investigated the effects of teacher codeswitching on vocabulary acquisition. According to him, the notion of *teaching English in English*, which is English-only instruction, is strongly supported by the Ministry of Education of Korea, although teachers find it difficult to follow this recommendation. Similarly, in the case of the present study, the Ministry of Education and Culture of Cyprus shows a similar position. The teacher's book, which has been prepared by the Ministry, states the following when referring to speaking activities: "The mother tongue should not be used in the classroom, otherwise it will interfere in the learning process. Only in exceptional cases the use of the mother tongue is recommended" (Georgiou, Kitromilides & Englezaki, 2007: 9). The use of codeswitching therefore, is clearly discouraged, without a valid justification.

2.3.3. Empirical studies

Empirical studies on classroom codeswitching can be divided into three main categories: a) teacher codeswitching studies, which either describe the codeswitching behaviour of the teachers or examine its effects on language learning; b) studies on the perceptions of teachers and students on classroom codeswitching; and c) studies on learner codeswitching. While all these types of empirical studies are discussed, greater space in this review is given to learner codeswitching studies, as this is the type of codeswitching that features in the present study.

2.3.3.1. Studies on teacher codeswitching

i. Descriptive studies

Earlier studies on teacher codeswitching adopted a descriptive approach, which focused on giving an account on the amount of L1 used by teachers and attempt to identify the functions that their L1 use serves. Looking at the amount of L1 use first, the findings demonstrate great variability. As shown in table 2.2.1 below, while in some contexts researchers report a percentage of L1 use as low as 0% (e.g. Duff & Polio, 1990; Rolin-Ianziti & Brownlie, 2002), in other contexts they report a higher percentage which could reach up to 90% or above (e.g. Duff &Polio, 1990; Giannikas, 2011). This inconsistency of findings is probably due to the variability between the studies, mainly in terms of research methods and the types of learners. This variability is illustrated by the information provided in the table below.

Table 2.3.1 Studies calculating the amount of teacher codeswitching

	Country	Languages	Number of teachers	Age of Learners	Method	L1 amount found
Duff & Polio (1990)	US	English (L1) 13 different L2s	13	University	Time analysis (15-seconds)	0-90%
Hosoda (2000)	Japan	Japanese (L1) English (L2)	1 *(Self-evaluation)*	Business school (Low proficiency)	Word count	18.74%
Macaro (2001)	UK	English (L1) French (L2)	6	Secondary school (11-14 year-olds)	Time-analysis (5-seconds)	4.8% (total lesson time) 6.9% (total talk)
Rolin-Ianziti & Brownlie (2002)	Australia	English (L1) French (L2)	4	University (beginners)	Word count	0-18.2%
Yanbin & Linjing (2002)	China	Chinese (L1) English (L2)	4	University	--------	1.3-5.6%
Liu, Ahn, Baek & Han (2004)	South Korea	Korean (L1) English (L2)	13	High schools	Word count	68% (average)
Kim & Elder (2005)	New Zealand	English (L1) 4 different FLs	7	Secondary schools	Counting AS units	12-77%
Edstrom (2006)	-----------	English (L1) Spanish (L2)	1 *(Self-evaluation)*	University (beginners)	Word count	0-71%
Kim & Elder (2008)	New Zealand	English (L1) French (L2) Korean (L2) [2 FL classrooms]	2	Secondary school	Counting AS units	French teacher: 12% Korean teacher: 77%
Raschka, Secombe & Chi-Ling (2009)	Taiwan	Mandarin (L1) English (L2)	2	*Cram schools* (secondary)	Counting instances of codeswitching	Teacher A: 268 instances Teacher B: 432 instances
Qian, Tian & Wang (2009)	China	Chinese (L1) English (L2)	2	Primary	Counting tag-switches, intra-sentential & inter-sentential codeswitching	Inter-sentential: 82% Intra-sentential: 16% Tag-switching: 2%
Giannikas (2011)	Greece	Greek (L1) English (L2)	7 state schools & 7 private schools	Primary state schools & private language centres	Timer (method not clear)	State schools: 60-98% Private schools: 0-50%

Concerning the type of learners participating, it is clear that their characteristics differ considerably. Looking at their age first, some are quite young (e.g. Qian *et al*., 2009), while others are adults (e.g. Duff & Polio, 1994; Rolin-Ianziti & Brownlie, 2002). The maturity of the learners plays a very important role, as more mature learners could be more intrinsically motivated than younger learners, and thus make a greater effort to understand what the teacher says in the L2. In such cases, the teacher is more likely to remain in the L2. The proficiency level of the learners also plays an important role, as younger learners are more likely to be of a lower proficiency level and thus require more switches to the L1 on the part of their teacher, in order to follow the lesson.

Moreover, the studies differ greatly in terms of the method they employ in order to calculate the amount of L1 and TL use. As shown in the table above, these methods include the word

count approach, counting instances of codeswitching or the time analysis approach, which notes what language is used every five (Macaro, 2001) or fifteen seconds (Duff & Polio, 1990). Due to these variations, it is hard to reach any generalisations from these studies. What can be reported as a tentative conclusion however, is that the L1 does exist in the FL classroom to some extent regardless of the context.

As for the functions of L1 use, research is more conclusive, as the findings are more consistent. As shown in table 2.3.2 below, two of the most frequent functions served by the L1 are classroom management and vocabulary translations. Classroom management, or lesson management, includes giving instructions on what is to happen in the lesson. The L1 is also frequently used for grammar instruction and solidarity, the latter being any attempt on the part of the teacher to build a relationship with the students and bridging the gap between them. Teachers also use the L1 for disciplining students, drawing their attention on specific forms, encouraging them to speak and for giving them feedback. Other less frequent functions are also listed in the table.

Table 2.3.2 Studies describing the functions of teacher codeswitching

	L1 FUNCTIONS																	
	Classroom management	**Vocabulary translation**	**Grammar instruction**	**Solidarity (incl. jokes)**	**Discipline/Controlling behaviour**	**Draw attention/ Highlight**	**Commenting on TL forms**	**Motivating/Encouraging**	**Feedback/Praise**	**Activity instructions**	**Check comprehension**	**Lack of comprehension**	**Providing background information**	**Explaining text/ideas**	**Demonstrating authority**	**Mark topic change**	**Assigning tasks**	**Informing about target culture**
Polio & Duff (1994)	✓	✓	✓	✓								✓						
Macaro (1997)	✓				✓				✓	✓	✓							
Tang (2000)	✓	✓	✓											✓				
Macaro (2001)					✓					✓								
Rolin-Ianziti & Brownlie (2002)	✓	✓					✓	✓										
Yanbin & Linjin (2002)		✓	✓			✓			✓			✓						
Liu, Ahn, Baek & Han (2004)	✓	✓	✓	✓	✓						✓		✓	✓				
Gearon (2006)						✓	✓						✓					
Kim & Elder (2008)			✓		✓					✓				✓				✓
Huerta-Macias & Kephart (2009)		✓	✓	✓														
Raschka, Secombe & Chi-Ling (2009)	✓			✓		✓	✓	✓								✓	✓	
Tien (2009)	✓	✓	✓	✓	✓										✓			
Qian, Tian & Wang (2009)		✓		✓	✓	✓		✓	✓									
Giannikas (2011)	✓		✓	✓														

Despite the different contexts of these studies, the results on L1 functions seem to be more consistent. Almost all of these studies report that the L1 can be used as an important

pedagogical tool that can serve a range of social functions. This wide range of functions, however, makes one wonder if there are any functions that the L1 is never used for. No study to date has attempted to answer this question.

ii. Experimental studies

More recent studies on teacher codeswitching adopted a more experimental approach in order to contribute to the construction of a framework that would establish the 'optimal' use of L1 in FL classrooms (Macaro, 2001). These studies, which in their majority were conducted at the University of Oxford, aimed to examine the relationship between codeswitching and one aspect of language learning, namely vocabulary acquisition.

In a study carried out in a secondary school in China, Meng (2005) attempted to examine whether teacher codeswitching led to better vocabulary acquisition. A hundred and thirty-nine participants were randomly assigned into two experimental groups, while an intact class acted as a third experimental group. While engaging in reading comprehension, the teacher provided L1 equivalents for the target words in the first group, L2 definitions for the second group and definitions plus L1 equivalents for the third group. Pre- and post-tests showed that these three approaches had very little or no effect on vocabulary acquisition. Nevertheless, what the results suggested was that teacher codeswitching did not have a detrimental effect on vocabulary learning, and since the codeswitching approach is more 'learner-friendly' then it should be preferred. The study however, has two important methodological problems. First, the lack of a baseline group does not allow the comparison of the effects of the intervention on the three experimental groups. Second, the third group received a lot more information than the other two groups (L2 definitions and L1 equivalents), making comparisons with the other two groups unequal.

Liu (2008) also examined the effect of teacher codeswitching on L2 vocabulary learning, again in an EFL context in a university in China. There were two groups in Liu's (2008) study: an experimental group, in which the teacher provided both English and Chinese explanations for unknown words and expressions; and a control group, in which the teacher explained unknown words in English only. As a pre- and post-test, the students were asked to translate English words and sentences into Chinese. Liu (2008) found that the experimental group did better than the control group. This should not be a surprising result however, since the experimental group received a lot more information and input on vocabulary, than the control group.

More recently, Tian (2009) conducted a two-month intervention study at a Chinese university in order to explore the impact of teacher codeswitching on the learners' L2 vocabulary learning. Participants, who were 117 first-year undergraduate English-major students, were encouraged to interact with the teacher in order to negotiate the meaning of new vocabulary in listening materials. Having randomly assigned the participants into three groups, this negotiation took place under three different conditions: for the *codeswitching group,* the teacher provided Chinese equivalents for L2 target words; for the *non-codeswitching group,* the teacher explained the target words in English; and for the *control group,* the teacher did not provide any information regarding the words, but the learners completed the same listening activities as the other groups. Pre- and post-tests were analysed in terms of vocabulary and listening comprehension. The results suggested that teacher codeswitching facilitated vocabulary learning. This effect however was only found to be short term.

Finally, Lee's (2010) large study in Korea provided more conclusive results. The study concerned the effect of teacher codeswitching on English vocabulary learning in two age groups: 12-year-old and 19-year-old students. Having adopted a quasi-experimental design,

two different kinds of intact classes participated in the study: an *English-only group*, in which the teachers were monolingual English speakers, and a *codeswitching group*, in which bilingual teachers used codeswitching for explaining unknown L2 vocabulary. With a large sample of 286 adult participants and 443 young participants, Lee (2010) found that the age groups under codeswitching conditions had more vocabulary gains than the age groups under the English-only conditions. In addition, younger learners under CS conditions benefitted even more than the adult learners.

As a conclusion to the experimental studies therefore, it could be argued that more recent studies have been able to support with empirical evidence that codeswitching may enhance L2 vocabulary acquisition (e.g. Tian, 2009; Lee, 2010).

2.3.3.2. Studies on teachers' and students' perceptions

The views of teachers and students on TL and L1 use in foreign language classrooms have been studied extensively over the years. One of the earliest studies on teachers' attitudes and beliefs towards L1 use in FL classrooms was the Tarclindy research project carried out by Macaro (1997). Having conducted surveys and interviews with teachers, Macaro (1997) concluded that teachers held three different theoretical positions towards the issue: the *virtual,* the *maximal* and the *optimal* position. The *virtual* and the *maximal* positions argued against the use of the L1 in the FL classroom, but they differed in the extent to which they implemented it. The *virtual* position suggested that the classroom should mirror the target country, with the TL being the sole medium of instruction. The *maximal* position argued that since perfect teaching and learning conditions did not exist, teachers would always need to resort to the L1. On the other hand, teacher supporters of the *optimal* position found some pedagogical value in the use of the L1. Macaro confirmed these positions in his 2001 study

with beginner teachers, whose beliefs also corresponded with one of these theoretical positions.

Another highly cited study on both teachers' and students' beliefs was an internet-based questionnaire survey conducted by Levine (2003). The aim of the survey was to give an account of what was going on in FL classrooms in terms of the distribution of TL and L1 use. This was a Likert scale questionnaire which was completed by 600 FL learners and 163 FL teachers from the USA and Canada. The responses of the teachers were compared to those of the students on each question. In terms of the amount of L1 use in the classroom, although the rates of the teachers differed to those of the students, the overall results suggested that the L1 was being used in the FL classrooms a great deal. Some indicative results were that 44% of the teacher respondents reported using the TL with their students 80% to 100% of the classroom time. Moreover, 53.3% of the respondents claimed that they used the TL 80% to 100% of the time for topic-based activities.

Although this study incorporated a large sample, the problem was the representativeness of that sample. The questionnaires were only completed by the teachers and students who had internet access and who were willing to complete the questionnaire. In terms of the students especially, it is likely that these were more motivated and proficient students, as they took the time to contribute to a survey which concerned language learning. A large part of the student community therefore is likely to have been omitted.

Crawford (2004) conducted a survey on language teachers' attitudes towards TL use in Queensland, Australia. The questionnaire was completed by 581 teachers who taught one of seven different foreign languages either in primary or secondary schools. This questionnaire aimed to examine the extent of desirability of the teachers to use the TL in the classroom, as well as to estimate the amount of L1 use in this context based on a Likert scale. The results

contradicted those of Levine (2003), who found that the L1 was being used in the classroom, but the TL was still the dominant language. The teachers in Crawford's (2004) study seemed to favour the use of the L1 as the main medium of instruction in most of their classes. They found this to be 'the appropriate medium for cross-lingual, cross-cultural comparisons' (Crawford, 2004: 5). In terms of the TL, the teachers seemed to have reservations about using it exclusively. In fact, the majority of the teachers reported using the TL for about 40% or less of the classroom time. Based on these results, Crawford (2004) concluded that 'current practices [in Australia] are not maximising students' exposure to the target language' (Crawford, 2004: 17).

More recent studies exploring the attitudes of learners on L1 use in the classroom found an overall positive response. Rolin-Ianziti and Varshney (2008) investigated the views of 52 university students in Australia. The beginner learners of French completed a questionnaire consisting of 21 closed questions, that were analysed quantitatively, and two open questions, that were analysed qualitatively. The results showed that the students attributed both negative and positive functions to L1 use. On the one hand, students found the L1 to be helpful in gaining explicit knowledge of TL linguistic features (e.g. vocabulary and grammar). They also acknowledged the affective role of the L1 in language learning, as a tool which alleviates negative feelings (e.g. pressure, frustration, confusion). On the other hand however, they also stressed the importance of being exposed to TL input. They thought that lack of TL exposure deprived them of learning 'naturally' from the context and acquiring pronunciation. According to the authors, therefore, an important insight that came from the results was that both the TL and the L1 were necessary for language learning.

Brooks-Lewis (2009) investigated the views of 256 Spanish-speaking adult learners in Mexico. Having had an unpleasant experience with teaching and learning a foreign language

herself, Brooks-Lewis (2009) designed a 30-hour introductory EFL course which incorporated extensive teaching in the L1 and gradual increase of the L2. At the beginning of the course therefore, the teacher used only Spanish and later on the teacher's use of English was increased. The data, which came from the students' anonymous written reports (diaries, essays, questionnaires), revealed that overall the perceptions of the learners towards the course were overwhelmingly positive. More specifically, the students reported that the presence of the L1 enabled them to participate in the classroom as they comprehended what was being said, it reduced their anxiety and it made learning easier as it allowed them to compare the new L2 knowledge. While there were a few students who expressed their preference for more L2 exposure from the beginning of the course, Brooks-Lewis (2009) thought that the incorporation of the L1 in L2 teaching was beneficial for the students.

Concluding on the perception studies therefore, it is clear that most of the teachers in the majority of the studies, not only admitted using the L1, but thought that this was sometimes necessary (Macaro *et al.,* 2009). It could be said therefore that the results of teacher perception studies confirm the results of studies on teacher codeswitching. Both types of studies provide evidence of the presence of the L1 in the classroom for a range of purposes. Although L1 use might differ in terms of the extent to which it is being used, the important conclusion here is that it is present regardless of context.

2.3.3.3. Studies on learner codeswitching

One of the main concerns of language teachers regarding student pair or group work is that students might see this as an opportunity to use their L1 (Storch & Aldosari, 2010). However, research conducted on learner codeswitching seems to suggest that although students use their L1 during pair work, this does not hinder their performance. In fact, it is used as a useful social and cognitive tool which is pedagogically valuable for L2 acquisition in both FL classrooms (Brooks & Donato, 1994; Antón & DiCamilla, 1999) and immersion classrooms (Swain & Lapkin, 2000). This section reviews only studies conducted in FL classrooms, as this is the focus of the present study.

As it will become evident when reviewing learner codeswitching studies, task-based learning and Vygotsky's *socio-cultural theory (SCT)* (1986) play a very significant role in these studies. Having already reviewed the literature on task-based learning, a section is dedicated here to the SCT framework, before the analysis of empirical studies. Following this section, empirical studies are presented in two separate sections: *studies with non-adult participants* and *studies with adult participants.* The rationale behind this distinction does not stem from potential proficiency differences, but rather from possible differences in maturity. Adult learners are more likely to make a larger effort to continue using the L2 when working in pairs, while younger learners could use pair work as an opportunity for excessive L1 use. Therefore, one is likely to find larger amounts of L1 use in studies with younger learners rather than in studies with adult learners. In addition to this, in a large study in Korea, Macaro and Lee (2013, forthcoming) found a difference between the perceptions and attitudes of two age groups of English learners. Younger learners welcomed the L1 in the classroom much more than adult learners. The authors argued that this difference between the two groups occurred due to the difference in terms of maturity, defined as strategic competence and learning experience, and not necessarily due to their difference in proficiency. For these

reasons, the maturity of participants is considered as an important variable, and this section is divided in such a way.

i. *Vygotsky's Sociocultural Theory (SCT)*

Socio-cultural theory (STC) emerged from Vygotsky's work in psychology in the 1930s. His theory concerned the relationship between language and thought and it assumed a very strong connection between the two. More specifically, Vygotsky perceived language to be the medium of thought, which allowed for higher mental activities, such as planning, organising and monitoring, to take place. Within this framework therefore, speaking is considered to be a highly cognitive activity, as it requires multiple cognitive processes to take place.

One of the central ideas of the SCT is that learning takes place through building a shared social reality. Meaning therefore is 'co-constructed through [this] social interaction' (Leeming, 2011: 362). One of the ways this sharing takes place is through *scaffolding*. Scaffolding refers to the assistance that one can provide to another in order to improve within the *zone of proximal development (ZPD)*. As Vygotsky (1986) argued when he referred to the learning experience of a child, the ZPD is 'the discrepancy between a child's actual mental age and the level he reaches in solving problems with assistance' (187). In terms of language learning, this refers to the distance between what learners can do on their own when solving tasks, and the level they can reach with the help of their teacher or of a more capable student.

Finally, another key construct deriving from Vygotsky's theory is *private speech.* This is an externalised kind of speech which is directed towards the speaker, and does not expect a reply from interlocutors. In Ohta's (2001) words, it is 'audible speech not adapted to an addressee' (16). It could come in the form of questions, instructions or evaluations of one's own performance (Ellis, 2008). Private speech occurs when a task is cognitively demanding and

causes stress (Ellis, 2008), in which case learners use it in order to gain control and self-regulation. A few SLA studies which examined private speech showed that it mainly serves two functions; it directs the learners' attention on a particular aspect of tasks, while it helps learners evaluate their ideas (e.g. Ohta, 2001; Centeno-Cortés & Jimémez-Jiménez, 2004).

Although the SCT is not a theory that directly explains language acquisition, it has been used extensively by SLA researchers in order to explain how learning occurs through building a social shared reality. Lantolf (2006) is one of the researchers that have shown through their work over the years how language learning can be seen from the SCT perspective. Within the field of learner codeswitching in particular, the SCT has been used extensively as a framework for analysing task language. This is because, in Leeming's (2011) words, it 'allow[s] a more extensive interpretation of students' language [...] incorporating rather than glossing over use of L1 and allowing for interpretation of the cultural context in which the interaction occurs' (362). Similarly, Vygotsky's framework is used in the present study for the analysis of task language during student interactions. Within this framework, instances of scaffolded help and private speech can be identified, while L1 use can be interpreted.

ii. *Studies with non-adult participants*

Researchers started being interested in learner codeswitching about two decades ago. A pioneer study conducted by Brooks and Donato (1994) re-analysed data of a previous study (Brooks, 1992) within a socio-cultural theory framework, in order to demonstrate that task language is a cognitive activity, rather than simply encoding and decoding verbal messages. Data were collected from the recordings of eight pairs of third-year English-speaking high school students, who completed a jigsaw task in Spanish. The main finding emerging from the qualitative analysis of the transcripts was that students' discourse did not just consist of

utterances which were part of the task, but of *metatalk* as well. Metatalk, which included comments on the words they should be using and decisions about how to complete the task, occurred in the L1. An explanation of why this occurred in the L1 was attributed to Vygotsky, who suggested that metatalk is the "out-loud" version of metacognition. Metacognition is constructed primarily through language, so it is not surprising that language based on metacognition is externalised in the L1. Concluding on their findings, Brooks and Donato (1994) reported that in their study the L1 was used as an important tool for '*initiating* and *sustaining* further discourse' (266). Therefore, they argued that the L1 should be allowed during L2 interactions, without necessarily being encouraged because 'it is a normal psychological process that facilitates L2 production' (Brooks & Donato, 1994: 268).

Eldridge (1996) reported on a small-scale study he conducted in one EFL classroom with 11 to 13-year-old students in a Turkish secondary school. Although Eldridge (1996) referred to the data collection methods he used (recordings of classroom lessons, noting, short student interviews), his description lacks basic information such as the size of the sample, the number of recording hours, as well as the way these data were analysed. Despite these problems, it is of interest that Eldridge (1996) found that overall the codeswitching behaviour of the students had a pedagogical purpose. Learners' switched to their L1 when they completed tasks (77%), when they commented on procedural matters or when they had questions (16%). Furthermore, the learners used the L1 for vocabulary explanations, to claim floor holding, for metalanguage, reiteration and so on. Based on these findings, Eldridge (1996) concluded that codeswitching is a strategy offering short-term benefits to the learner 'but with a risk of hampering long-term acquisition' (310) when it is overused.

Another study which focused on learner codeswitching during group work was conducted by Hancock (1997). Thirty six Spanish EFL students, aged 14 to 17, participated in the study.

Data were collected from recording the participants' interactions while completing tasks. The participants in pairs completed two tasks with a two-week interval. For the first task, a restaurant role play, no planning time was given, but the students had the opportunity to choose their interlocutor. For the second task, a guesthouse roleplay, a few minutes were provided for planning, but the students were paired with students from another class, who they were unfamiliar with. 'Planning time' and 'interlocutor familiarity' therefore, were two of the variables under consideration. It would be reasonable to assume, thus, that more L1 would occur in the restaurant roleplay for two reasons: firstly, because the speakers had no planning time so they would need to talk about how to complete the task, a type of discourse which normally occurs in the L1; and secondly, because they would feel more comfortable with a person they had chosen and thus switch to the L1 more often.

Using Goffman's (1974) framework to analyse the data, Hancock (1997) distinguished between two layers of discourse: the *literal frame*, that is the real world outside the task in which *off-record* discourse occurs; and the *non-literal frame*, that is the world of the task in which *on-record* discourse occurs as part of their role. Hancock (1997) found that off-record L1 could be divided into three main categories: *metatask language,* which establishes who should say what and when (e.g. turn-dispute, prompts); *metalanguage,* which includes modelling (providing a translation), or a translation appeal (asking for a translation); and less often *self-address*, which is talking to one's self. In terms of *on-record* discourse, the L1 occurred for turns-in-role, insertions and jokes.

In terms of the tasks, Hancock (1997) found that, as expected, the guesthouse roleplay had fewer occurrences of on-record Spanish than the restaurant roleplay. Hancock attributed this result to the interlocutor variable, saying that the speakers were 'less drawn to converge to the (unknown) interlocutor' (Hancock, 1997: 232). Topic familiarity emerged as another

influential variable, as the students were more familiar with the topic of the restaurant roleplay. This familiarity inspired more involvement which perhaps led to more unintentional slips into the L1. No reference was made to the 'planning time' variable. Hancock (1997) concluded that 'the design and setup of the task will affect the quality of group work' (233). Finally, although 'planning time' was not discussed in the results of the paper, the fact that more L1 was used in the restaurant roleplay partly confirmed that no planning time results in more L1 use.

A more recent study which analyses students' speech during group work was conducted by Alley (2005). While the overall aim of this study was to examine the advantages and disadvantages of group work, and not specifically L1 use, the paper has been chosen for discussion because of the interesting categorisation of L1 use. Eighteen English-speaking high school students who learned Spanish participated in the study. The students were recorded over a period of four months while they completed a series of different tasks in small groups (e.g. interviews, role-plays, narrative tasks). Having transcribed the students' conversations, the researcher categorised all non-task discourse into one of the following three categories: *metatalk,* which included discourse about their own talk; *metacognitive talk,* which included talking about how to do a task; and *off-task talk,* that is talk which is irrelevant to the task. Each category was then further analysed into L1 and L2 instances, in order to compare the number of times English was used to the number of times Spanish was used. However, this method is somewhat problematic. Counting instances of codeswitching makes the categorisation of intra-sentential codeswitching very difficult. It comes down to the researcher to judge whether such utterances are more L1 utterances or L2 utterances, in which case the method becomes unreliable.

Nonetheless, Alley (2005) reported that 48% of the discourse was for metatalk, 31% for metacognitive talk and 21% for off-task talk. Regarding distributions within each category, more than half of metatalk occurred in the L1 (57%), while all metacognitive talk (100%) and off-task talk (94%) occurred in the L1. Overall, most of the group work discourse occurred in the L1 (71%). Nevertheless, Alley (2005) concluded that, although group work produced a lot of L1 and off-task talk, it 'improve[d] the quality of language practice' (256) because, under these circumstances, students were able to recognise each others' problems more readily than teachers and help each other. In terms of the role of codeswitching in the study, Alley (2005) argued that the students' use of English served to clarify task procedures and to produce key vocabulary items.

The work of Carless (2002, 2008) was already discussed in the section on teachers' opinions on task implementation. However, as his studies in classrooms in Hong Kong included observations and his own arguments on learner codeswitching, his work should be discussed from another perspective in this section as well. In a study in primary schools in Hong Kong, Carless (2002) found that students used Cantonese more than English during tasks. At times this L1 use was relevant to the task, but at other points it was off-task talk. Carless (2002) commented therefore that while L1 use served as a social function, it is unlikely that it contributed to language learning. Two variables were found to be related to the amount of L1 use: task complexity and students' proficiency level. The more complex and open-ended tasks were, and the less proficient students were, the more L1 was used. Furthermore, in another study in the same context, Carless (2008) claimed that Cantonese high school learners used the L1 widely and this could undermine the task-based learning approach. However, the problem with Carless' studies is that they lack crucial information. We know for example, that he observed these students for a long period of time, but we do not know how he determined the amount of language, the functions they served or the complexity of tasks.

More recently, Leeming (2011) conducted a study examining the codeswitching behaviour of Japanese high school learners of English during pair work. Two pairs of female students completed two-way information gap tasks, while being audio-recorded. These tasks were part of the regular syllabus and they were performed under normal classroom conditions; this is rarely the case in learner codeswitching studies. In terms of the amount of L1 use, the percentage of language in Japanese was 38% for the first pair and 8% for the second pair. A qualitative analysis of their use of Japanese revealed that this was used for inter-psychological speech (e.g. scaffolded help, intersubjectivity), as well as intra-psychological speech (e.g. private speech); results which coincide with those of Antón and DiCamilla (1999), which is discussed later on in this section. Nevertheless, Leeming (2011) also found that Japanese was used for maintenance of the fluency of communication, mainly achieved by replacing an unknown English word with the Japanese equivalent. As for the attitudes of the students, interviews revealed that the students found that their L1 use had positive effects on their L2 learning. They believed that the presence of the L1 in the classroom created a positive classroom atmosphere, whereas an English-only environment would not have allowed them to speak.

iii. *<u>Studies with adult participants</u>*

One of the most cited studies on the codeswitching behaviour of adult learners was that of Antón and DiCamilla (1999). Following the work of the Brooks and Donato (1994) study, this study was also conducted within Vygotsky's (1986) SCT framework. The aim was to examine how various features of the L1 play a strategic role in student collaboration during task completion. Five pairs of adult English-speaking learners of Spanish participated in the study. These were beginner learners who attended a six-week intensive Spanish class. Data were

collected by recording the students' interactions while they were engaged in three writing tasks.

Having transcribed the data, Antón and DiCamilla (1999) found that the L1 had three important functions. First, it allowed students to provide *scaffolded help* which, as explained earlier, is the help provided from an expert to a novice, while they are engaged in a problem-solving task. The expert takes control of the parts of the task, which are beyond the novice's current level of competence, while the learner focuses on the elements which are within his/her ability. Examples of L1 used as scaffolded help are accessing L2 linguistic items, reflecting on their own language and understanding the meaning of a text. Second, the L1 helped to achieve Rommetveit's (1985) *intersubjectivity;* that is a shared understanding of the goals, objects and events of the task. Such L1 uses included controlling the task, limiting the goals throughout the task and making the task manageable. Third, using the L1 helped these learners to externalise their inner speech, namely *private speech,* which reflected the cognitive processing the learners went through. Examples of private speech found in the transcripts include self-evaluations of what has just been produced and vocalising and immediately answering questions.

Based on the above findings, Antón and DiCamilla (1999) concluded that the L1 is beneficial because it has a facilitative role in language learning. More specifically, they argued that it 'acts as a critical psychological tool that enables learners to construct effective collaborative dialogue [...]' (245). As an implication, the authors note that, since language and thought are bound together within the SCT framework, prohibiting the L1 use in the classroom removes 'two powerful tools for learning: the L1 and effective collaboration' (Antón & DiCamilla, 1999: 245). However, the functions of the L2 were not examined. Perhaps scaffolding, intersubjectivity and private speech would have occurred under L2-only conditions as well. In

other words, there is no clear evidence suggesting that the L1 is the only tool that is able to serve these functions. An examination into the functions served by the L2 might have shown that the learners are able to perform these functions in the L2 as well, when they face the challenge of L2-only conditions. The conclusions reached by the authors regarding the importance of the L1 in task completion therefore seem to be somewhat one-sided.

Fotos (2001) observed Japanese university students learners of English, in order to examine their L1 use during grammar tasks. She found that Japanese was used as a tool for conversation when they sought clarifications, when stressing a point and for repair work. Similarly, Storch and Wigglesworth (2003) investigated the use of L1 of three Indonesian and three Chinese university students learning English in a second language context. The students, who were working in pairs, completed a text reconstruction task and a joint composition task. The researchers concluded that task type affected the functions served by L1 use. The L1 was used more for task management and clarifications during the joint composition task, while it allowed discussions on vocabulary and meaning during the reconstruction task. As for the amount of L1 use, this varied across the three pairs. Student interviews showed that while students were hesitant to use the L1, they found it to be helpful in completing the tasks.

In a more recent study, de la Colina and Mayo (2009) analyse the L1 use of Spanish undergraduates, who were low proficiency English learners. The participants formed twelve pairs and completed three different collaborative tasks. Three pairs worked on a jigsaw task, three on a text reconstruction of a written stimulus, and three on a dictogloss, that is a text reconstruction of a listening stimulus. Having recorded and transcribed their conversations, the authors found that the L1 was used for two main purposes: *metacognitive talk* and *metatalk.* Metacognitive talk involved task management, that is planning, organising, monitoring the activity, setting goals and checking understanding. Metatalk included episodes

of discussing vocabulary and grammar. Some *off-task* talk also occurred and it involved casual talk which was unrelated to the task. Adopting a word count method, the researchers reported a high percentage of L1 use on all three tasks (77% in text reconstruction, 75% in dictogloss, 55% in jigsaw).

De la Colina and Mayo (2009) concluded that the use of the L1 helped learners in their study perform higher level activities and, thus, enhance their motivation. However, as in the Antón and DiCamilla (1999) study, the results do not support this conclusion. What the results actually show is a facilitative role of L1 use, which helps the completion of tasks. They do not show that the cognitive level in these activities was higher than it would have been under L2-only conditions, simply because no such conditions were tested. Perhaps a comparison of these data with a group of learners who completed the same tasks under L2-only conditions would have provided a better foundation for supporting such a conclusion.

Another study examining pair work at a college in Saudi Arabia was conducted by Storch and Aldosari (2010). This study was concerned with the potential effect of proficiency pairing and task type on the amount and functions of L1 use (Arabic). Thirty male students from two EFL classes formed fifteen pairs. Five pairs consisted of two high proficiency students (HH), five consisted of one high and one low proficiency student (HL) and the latter five pairs consisted of two low proficiency students (LL). Their level of proficiency had been determined by the teacher, who based his opinion on test scores and in-class performance. All pairs were asked to complete a total of three tasks, namely a jigsaw task, a composition task and a text-editing task, while being audio-recorded. It should be noted that these students were not familiar with completing tasks, as the language teaching environment in that context was more traditional, involving teacher-centred instruction and no communicative activities.

Having analysed their interactions in terms of L1 words and L1 turns, Storch and Aldosari (2010) found that overall the students were very careful in their L1 use, with just 7% of their total speech occurring in the L1. In terms of proficiency pairing, the LL pairs switched to the L1 more frequently (12%) than the other two proficiency groups (5%). In terms of the task type effect, it was found that the LL pairs used a lot more L1 words (29%) on the editing-task, in comparison to the other tasks. As for L1 functions, Storch and Aldosari (2010) found that the L1 was used mainly for task management. It was also used however for explaining vocabulary and for private speech. The LL pairs additionally used the L1 for negotiation of grammar, spelling and punctuation. The HH pairs used it for generating and discussing ideas.

Based on these results therefore, the authors concluded that the L1 served 'important cognitive, social and pedagogical functions' (Storch & Aldosari, 2010: 372), while the amount of L1 seemed to be more affected by the type of task, rather than proficiency pairing. Nevertheless, since the study did not look at functions of L2 use, then the argument supporting the L1 as an important tool because it serves these functions, is not convincing. If the study proved in some way that the L1 is the only tool able to serve these functions, then their conclusion would have been justified. But by ignoring the possibility of these functions being served by the L2 as well, then such an argument cannot be made.

A very recent study on learner codeswitching is a small-scale study conducted by Macaro *et al.* (2012). Forty-nine Japanese undergraduate students, who attended a three-week intensive EFL course in the UK, participated in the study. These students were taught by a teacher who was a monolingual speaker of English (i.e. non-speaker of Japanese), as well as a teaching assistant. Having divided the participants into three groups based on C-test scores and productive vocabulary scores, the researchers placed a monolingual assistant in the first two groups and a bilingual assistant of Japanese and English in the third group. The 'bilingual'

group was encouraged to consult this assistant regarding the meaning of L2 words, as well as the L2 equivalents of L1 phrases. More specifically, this group was actively encouraged to codeswitch, when they faced difficulty in expressing themselves in the L2, while the first two groups were discouraged from using the L1 among themselves. Data on students' perceptions were collected through an open-ended questionnaire, while data on the impact of the 'intervention' were collected with an oral production test which acted as a pre- and post-test.

In terms of the L2-only conditions, the students in general had a positive attitude towards this because they found it challenging. Others however commented that this 'reduced the length and complexity of [their] output' (Macaro *et al.*, 2012: 7). As for the bilingual group, the majority of the students had a generally positive attitude towards allowing the L1 in the classroom. Their responses reflected a real benefit in terms of listening skills, oral production skills and the use of higher cognitive level. In addition, although the results were not statistically supported, the authors argued that the 'bilingual' group showed the largest improvement in terms of fluency and overall speaking scores.

Although this is a small-scale study which did not yield statistically significant results, it is innovative in the sense that it explored an original idea: allowing codeswitching may increase willingness to communicate and consequently oral production. It is a clear initial attempt therefore to contribute to the construction of the very-much needed framework that will indicate when codeswitching is beneficial for language learning and when it is simply the easy option (Macaro, 2001).

A different approach to the study of L1 use in the classroom was adopted by Moore (in press). Twelve Japanese undergraduate students, who were intermediate-level EFL learners, participated in the study by attending an oral presentation class once a week. Data coming from their interactions, while completing tasks, were analysed in terms of the amount and

functions of L1 use, as well as 'contextual features surrounding the emergence of L1 in EFL task-based interaction, over time and across participants' (Moore, in press: 9). In other words, Moore wanted to investigate how and why L1 use occurs, rather than simply describing it. He found a high overall proportion of L1 use (28%). This was reduced across time and not always affected when the interlocutor changed. In terms of the contextual surroundings of L1 use, Moore argued that these could be divided into four categories: *procedural* (e.g. interpreting the rubric), commenting on *performance,* creating *content* and for *off-task* talk. Based on his findings, Moore concluded that teachers and learners should develop an awareness of how the L1 is naturally occurring in the classroom. This awareness could then inform task design and task implementation.

Finally, DiCamilla and Antón (2012) conducted a study which aimed to identify the role of proficiency level in the functions served by the L1 and L2 during a writing task. Two groups of a total of twenty-two English-speaking university learners of Spanish participated in the study: an advanced-level group and a beginner-level group. According to the authors, this study built on previous research, as this comparison of linguistic behaviour between the two groups of students provides a better 'understanding of the mediating function of language in general (L1 and/or L2)' (DiCamilla & Antón, 2012: 161). Having recorded the interactions between the learners, two different approaches were adopted for the analysis of the data: a quantitative approach, which involved counting the number of words in the L1 and the L2; and a qualitative approach, which involved analysing language within Vygotsky's SCT framework.

The quantitative analysis determined that the beginner group used a lot more L1 (70%-82%) than the advanced group (0%-3%); a result which according to the authors was due to proficiency differences. Beginners used the L1 mainly for creating content, for solving

problems and for interpersonal relations. As these were integral parts for task completion, the authors characterised this use as 'the primary meditational device for performing the task' (183) and commented that if beginners were deprived of using their L1, then they would probably not be able to complete the task at all. Beginners regarded the L2 as the system that should be learned, rather than the system that should be used for learning (DiCamilla & Antón, 2012). As for the advanced learners, their limited L1 was used for unknown L2 words, for marking breaking points of the task and for externalising private thinking. Their use of the L2 served functions, such as creating content, solving problems and interpersonal relations; the same functions that beginner learners used their L1 for. For advanced learners, the L2 was the system they used in order to learn, rather than 'the system to be learned' (DiCamilla & Antón, 2012: 184).

Based on these findings, DiCamilla and Antón (2012) concluded that their study provided important evidence supporting the view that the L1 has 'communicative, cognitive, and, hence, real pedagogical value' (185) in L2 learning. Nevertheless, what distinguishes this study from previous studies is that it is the only one that looks at the functions of both the L1, as well as the L2. It proves that as proficiency increases, the same functions served by the L1 can be served by the L2. Despite the conclusion of the authors referring to the importance of the L1 in language learning, what should be reported is the fact that learners are able to use the L2 for metalanguage. This idea is explored further in the present study.

iv. Conclusions on learner codeswitching studies

Concluding on learner codeswitching studies, it seems that most studies in this area are small-scale and of a more exploratory nature. The methodology adopted in these studies is quite similar. In most of them, participants formed pairs and completed certain tasks while being

recorded in laboratory settings, which may have an effect on the learners' performance. As Foster (1998) claims, it should not be assumed that 'the style adopted by learners in an undisturbed classroom will be the same as that adopted by learners in an experimental set-up' (4). In an attempt to avoid the effect of laboratory environments, the present study was conducted in a natural classroom setting.

In terms of the results of these studies, it seems that researchers focused their attention on the amount and functions of L1 use, as in the teachers' studies. Concerning the amount of L1 use, this seems to vary according to context and not according to maturity as it was predicted. While some studies reported a very low percentage of L1 use (e.g. Storch & Aldosari, 2010), others reported a very high percentage of L1 use (e.g. Storch & Wigglesworth, 2003; Alley, 2005). As for the functions of L1 use, most researchers concluded that the L1 was used as a useful social, cognitive and pedagogical tool in their studies (e.g. Storch & Aldosari, 2010), which facilitated collaborative work (e.g. Antón & DiCamilla, 1999). Within the SCT framework, the L1 was often used for cognitive functions, such as metatalk, intersubjectivity and private speech. These studies concluded therefore that the L1 facilitates L2 acquisition, as it provides additional cognitive support to the learners. Only very few researchers expressed their concern regarding learners' L1 use being a barrier for language learning (e.g. Eldridge, 1996; Carless, 2002).

None of these studies however, seems to be exploring the potential benefits of an L2-only approach. Forced output could have a real impact on task performance, as it challenges learners more and does not allow them to abandon their effort to remain in the L2, when a difficulty arises. It can be argued therefore that there is an imbalance of studies in the area of learner codeswitching, as these studies seem to be taking for granted that the L1 will occur and therefore its role should be examined. Perhaps if there was an equal amount of empirical

studies analysing task performance under L2-only conditions, then an argument in favour of the facilitative role of L1 use would be more valid. Forced output conditions thus are tested as part of the present study.

In conclusion, researchers seem to argue that learners should be allowed to use their L1 during L2 task interactions. However, simply allowing students to codeswitch while completing tasks does not meet the aim of task-based learning, which is effective L2 communication. Learners should gradually learn how to move away from this state of reliance on the L1 to a state where they are able to communicate effectively in the L2 both for task language and for metalanguage. Although metalanguage is part of metacognition which, in Vygotsky's terms, is interwoven with language, namely the L1, this does not mean that metalanguage can only be expressed in the L1. As it was suggested by DiCamilla and Antón (2012) in their latest study, both the L1 and the L2 can be used for the same functions; the choice of language lies more on the proficiency level of the learners. Taking this a step further therefore, it can be argued that metalanguage can be expressed in the L2 as well, when a learner possesses the required L2 knowledge. The present study is an attempt to address this issue empirically.

2.3.4 Conclusions on codeswitching

Although classroom codeswitching was a "taboo" issue in language teaching methodology in the 1980s and the 1990s, an overwhelming amount of research has been conducted on this phenomenon, mostly within the past two decades. Empirical studies on classroom codeswitching have focused their attention on three different areas. Concerning studies on teacher codeswitching, descriptive studies provide evidence that, regardless of the context and the type of learners, teachers use the L1 to some extent in order to serve a range of social or

pedagogical functions. Experimental studies on teacher codeswitching provide evidence suggesting that codeswitching facilitates vocabulary acquisition.

The result of descriptive studies on teacher codeswitching is also confirmed by the results of studies on teachers' and students' perceptions on classroom L1 use. Despite the variability in the reported amount of L1 use, these populations confirmed the presence of the L1 in different FL contexts. They also showed their approval towards it, as it serves a number of functions that facilitate the lesson.

Studies on learner codeswitching have shown that learners tend to use the L1 when working in pairs or small groups to various extents, but this use serves a range of social, cognitive and pedagogical functions. Within the SCT framework, the L1 was found to serve various cognitive functions (metalanguage, intersubjectivity, private speech). Researchers therefore have argued that L1 should be allowed during L2 interaction. In an attempt to take this type of research a step further however, it has been proposed in the present study that it should not be enough to simply allow learners to codeswitch because this does not serve the ultimate purposes of task-based learning. It is argued therefore, and is incorporated in the present study, that something additional should take place in order to move learners to a state of less reliance on their L1. There is no theory suggesting that using metalanguage in the L2 could not provide the same cognitive support as L1 metalanguage.

2.4. Oral fluency

2.4.1 Defining fluency

Before discussing oral fluency research, it is crucial to define what fluency is exactly. In a review of the origins of the word "fluency" in various languages, Kaponen and Riggenbach (2000) determined that the meaning of the word has an underlying conceptual metaphor, that of 'language in motion' (7). As Segalowitz (2010) argues however, if people are asked about what they perceive as fluent L2 speech, a range of different behaviours could be mentioned. Some might perceive this to be the ability to speak with an L2 accent; others might think that it has to do with the speed of production or even the range of vocabulary used. The concept of fluency therefore is often confused with that of oral production.

Researchers have adopted either a broad sense of fluency or a narrow sense (Lennon, 1990). In the broad sense, oral fluency is the general oral proficiency, or as Lennon (1990) puts it, 'the spoken command of a foreign language' (389). In the narrow sense however, fluency is one isolated component of oral production which has to do with 'smoothness of performance' (Segalowitz, 2007: 181). More recently, researchers seem to adopt the narrower sense, since there is consensus that oral proficiency consists of three components: accuracy, complexity and fluency. The narrow sense of fluency is also adopted in the present study, since this will be dealt as a purely performance phenomenon.

2.4.2. Measuring fluency

The present study is largely concerned with measuring oral fluency. A number of different ways of measuring fluency have been proposed in the past. Research on fluency began with the pioneer work of Frieda Goldman-Eisler (1951, 1968, 1972), who investigated the fluency

of native speakers for the first time. Research on the fluency of L2 speakers started in the 1970s and 1980s (Segalowitz, 2010). Empirical studies on L2 fluency investigated the development of fluency in three different ways (Kormos, 2006). Some adopted longitudinal designs and looked at how fluency developed over time (e.g. Lennon, 1990; Towell *et al.,* 1996); others compared the fluency of native speakers to that of non-native speakers (e.g. Riggenbach, 1991; Temple, 2000); while others have attempted to find whether there is a correlation between fluency measures and temporal variables (e.g. Fulcher, 1996).

According to Lennon (1990), oral fluency has two components: a *temporal component,* which has to do with time measurements, and a *dysfluency marker component,* which relates to smoothness of performance. The temporal component has been measured by calculating *speech rate*, that is words per minute (e.g. Lennon, 1990) or syllables per minute (e.g. Towell *et al.,* 1996; Ahmadian, 2011; Ahmadian & Tavakoli, 2011) or syllables per second (Temple, 2000; Iwashita *et al.,* 2008; Kessler, 2010). Similar to the speech rate measure is the *articulation rate,* which is the total number of syllables divided by time, but excluding pause time (Kormos, 2006). *Mean length of runs (MLR),* which is the average number of syllables produced in utterances between short pauses, has also been used as a temporal measure before (Lennon, 1990; Towell *et al.*, 1996; Kessler, 2010). However, it should be mentioned here that, although researchers conceptualise MLR as a temporal variable, one could conceive it as a dysfluency marker because it is related to length of pauses. Finally, the temporal component has been measured with the *phonation-time ratio,* which is the ratio of the percentage of time spent speaking to the time taken to produce the speech sample (e.g. Lennon, 1990; Towell *et al.* 1996).

Pause measurements have also been used in the past as indicators of fluency. *Pause rate* has been used as the proportion of filled and unfilled pauses (Lennon, 1990; Temple, 2000). Also,

the *silent ratio* calculated the percentage of speaking time spent in silent pauses (Temple, 2000). Towell *et al.* (1996) added *average length of pauses* in their fluency measures. Finally, Kessler (2010), who studied the relationship between fluency and anxiety of forty university students in America, preferred counting the amount of pauses that exceed 3 seconds per minute. Similarly, Bygate (2001) counted the number of unfilled pauses per *t*-unit, as the sole measure of fluency in his study.

Concerning measurements of the dysfluency marker component, researchers have calculated the number of *repetitions* (Lennon, 1990), *reformulations,* which are phrases repeated with some modification (Lennon, 1990; Foster & Skehan, 1999), and *false starts* (Ellis, 2008).

Lennon's (1990) widely cited study, which initiated the latest wave of research on L2 fluency, aimed to identify a set of measures that would be the best indicators of fluency. Four female university students from Germany, who spent six months at the University of Reading in England, participated in the study voluntarily. They were recorded narrating the same story shown on a sequence of six pictures twice; once upon their arrival at Reading and once six months later, shortly before their departure. A panel of native speaker EFL teachers judged their performance based on a set of twelve measures, including speech rate, mean length of runs, as well as a set of pause measures and repair measures. With the panel's agreement that the fluency of the students had improved, Lennon concluded that the reduction of filled pauses and repetitions, the improvement of speech rate and the reduction of pauses were the measures that contributed to the increase of perceived fluency.

It is clear therefore that a number of fluency measurements have been proposed in the past and researchers have used different combinations of these variables. Their choice of measures depended on the way each one of them perceived fluency. It should be noted however, that the number of participants in these studies has been relatively small (Kormos, 2006), making

it hard therefore to reach any generalisations. Despite the variation in measurement choices and the small sample size, the majority of these studies found that speech rate and mean length of runs were the best predictors of fluency (Kormos, 2006).

2.4.3. Fluency development

Another strand of research on fluency looked at how oral fluency develops. Reference to two frameworks is crucial for answering this question. First, Levelt's (1989) *Model of Speech Production* (Figure 2.4.1 below), which was designed to represent the language production procedure of monolingual adult native speakers, has been adapted for L2 speech production by several researchers (de Bot, 1992; Towell, Hawkins & Bazergui, 1996). Levelt's (1989) model consists of three processing components: the *conceptualiser,* the *formulator* and the *articulator.* In the conceptualiser, the intended message is encoded in a preverbal form. This preverbal message acquires linguistic structure in the formulator, by accessing information on meaning, syntax, phonology and morphology from the lexicon. These processes result in a phonetic plan, which is later turned into speech production by the articulator.

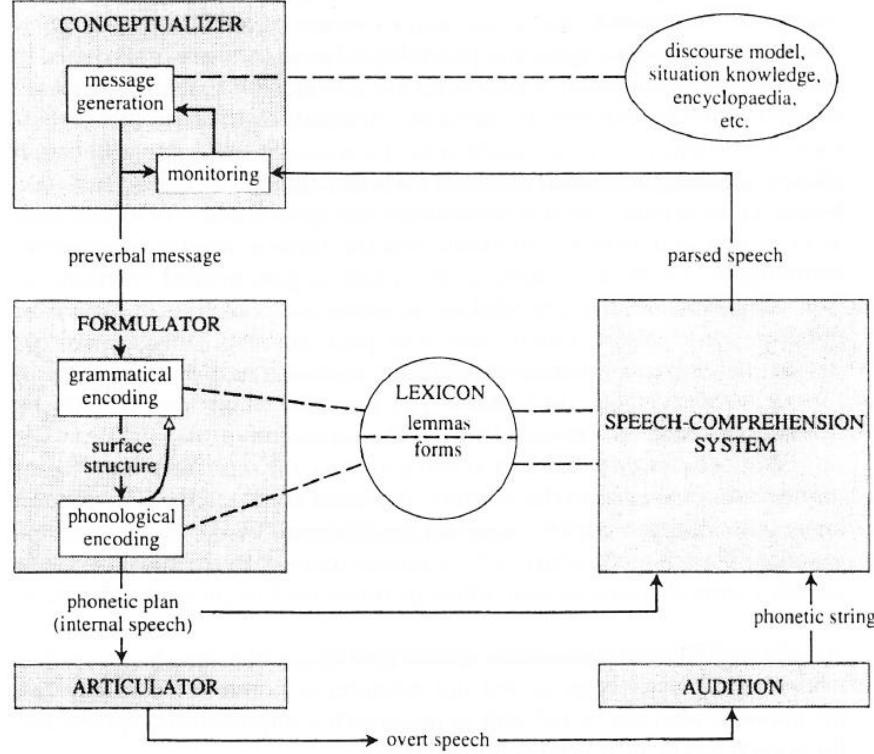


Figure 2.4.1 Levelt's Model of Speech Production (Levelt, 1989: 9)

For encoding messages, two types of long-term knowledge have to be accessed: *declarative knowledge* and *procedural knowledge.* Declarative knowledge, represented by the circles in Levelt's model, is the 'speaker's structured knowledge of the world and himself, built up in the course of a lifetime' (Levelt, 1989: 10). This is a more conscious kind of knowledge that requires more time for processing (Segalowitz, 2010). Procedural knowledge however, represented by the square boxed, is a more unconscious and automatic kind of knowledge. According to Levelt (1989), fluent speech requires proceduralised knowledge in every component of the model. This 'unthinking behaviour' (Towell *et al.,* 1996: 89) can deal with the demanding speed requirements of having a fluent conversation.

Levelt's model however does not explain how this proceduralised knowledge occurs. In order to fill this gap, Towell *et al.* (1996) used Anderson's (1983) *Adaptive Control of Thought* model. This model consists of three types of memories (Figure 2.4.2): two long-term memories, the *declarative* and *production memory* described above, which are connected to

the outside world through a third temporary memory store of limited capacity, namely the *working memory.* A range of processes take place between these memory stores. *Encoding* processes transfer information from the outside world to the working memory, while '*performance* processes convert commands in working memory into behaviour' (Anderson, 1983: 20). *Storage* processes convert the contents of the working memory into declarative knowledge, while *retrieval* processes retrieve information from the declarative memory. *Match* processes allow data to correspond with the conditions of production, which are then transferred into the working memory with the *execution* processes.

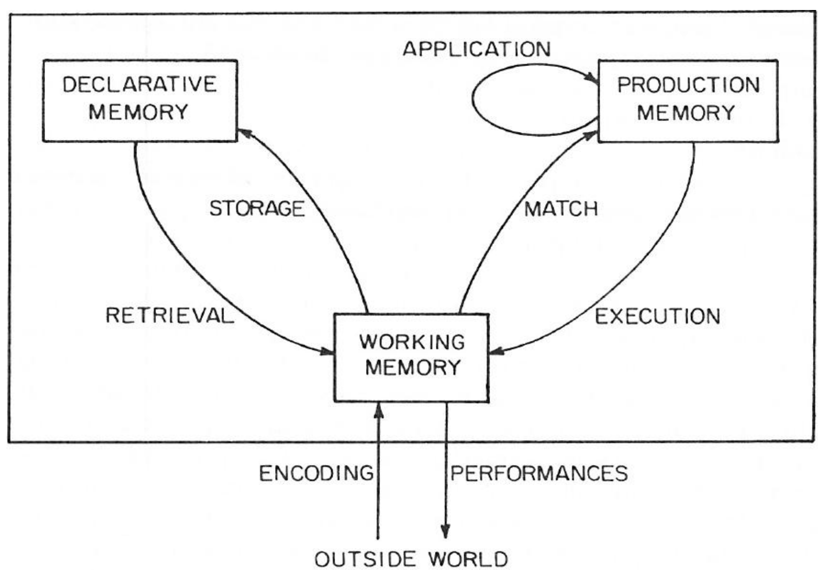


Figure 2.4.2. Anderson's Adaptive Control of Thought Model (J.R. Anderson, 1983: 19)

According to this model, all knowledge is initially declarative (Anderson, 1983). For fluent speech to occur however, declarative knowledge should be converted into production memory through a *proceduralisation* process. Proceduralised knowledge takes up much less 'space' in working memory and can therefore be processed much faster by the working memory, meeting the speed requirements of speech production. This conversion happens in three stages: the *cognitive stage,* the *associative stage* and the *autonomous stage* (Anderson, 1983). In the cognitive stage, knowledge is declarative and access to it is very slow. In the

associative stage, where both types of knowledge exist, processes are faster, but they are still slowed down by the need for declarative knowledge. In the autonomous stage, knowledge is procedural only and can be rapidly accessed.

But the question is how this conversion from declarative to proceduralised knowledge takes place. According to Anderson (1983) whose model refers to learning in general, proceduralisation occurs 'through practice' (34). This has been confirmed in terms of oral fluency development specifically by Towell *et al.* (1996) who investigated how declarative knowledge is converted to procedural knowledge. From their study, they concluded that 'as exposure to and practice of L2 increases so does fluency' (98). An example of the differences before and after the conversion is given by Kormos (2006), who explains that fluent L2 speakers who once knew the rules of using a tense, for instance the present-perfect tense, may now not remember what these rules are; this rule application process becomes automatic.

Together with another process, namely *composition,* proceduralisation is called *knowledge compilation* (Anderson, 1983). Composition involves the creation of macro-production, or in other words chunking, by combining 'smaller units of processing' (Kormos, 2006: 41). Proceduralisation and composition are two of the five learning mechanisms that help develop automatic performance. The other three are *generalisation,* which applies the declarative knowledge required for production to a wider scope of appropriate contexts; *discrimination,* which narrows the scope of application to appropriate contexts only; and *strengthening,* which weakens poorer rules and strengthens the ones that are more frequently used. The present study focuses on proceduralisation processes only, as these are the ones required for fluent speech production.

The study conducted by Towell *et al.* (1996) was the first study to have used Levelt's (1989) model of speech production and Anderson's (1983) Adaptive Control of Thought model as

frameworks for investigating the way fluency develops. This was a four-year longitudinal study which aimed to examine how knowledge from multiple sources contributed to fluent speech. Twelve English-speaking university students, who were advanced learners of French, participated in the study. These students were recorded while narrating the story of a short film three times: once in French on their second year; another time in French on their third year, after spending a year abroad in a French-speaking country; and once in English on their fourth year. Speaking rate, phonation/time ratio, articulation rate, mean length of runs (MLR) and average length of pauses were used as measures of fluency.

A quantitative analysis of the data showed that the fluency of the students improved before and after their year abroad in terms of speech rate, articulation rate and MLR. Their L2 fluency however, was still at a lower level in comparison to their L1 fluency. Answering to the question of what becomes proceduralised, Towell *et al.* (1996) qualitatively analysed transcripts from two students. They found that the increase of fluency was the result of the increase of the MLR, determined by longer and more complex structures, rather than the reduction of pausing or increase of speed. They perceived that as evidence of proceduralisation therefore and declared MLR as the best indicator of fluency.

2.4.4. Conclusions on fluency

Due to the complicated nature of the construct of fluency, researchers have proposed various ways in which this can be measured. Although researchers seem to agree on certain variables that appear to be the best predictors of fluency, such as speech rate and MLR, they should first define what they perceive fluency to be and then set the appropriate measures. The other issue is the development of fluency. As in the Towell *et al.* (1996) study, Levelt's (1989) and Anderson's (1985) models were adopted in the present study, in order to examine how

fluency develops. As suggested by the models, extensive oral practice leads to the necessary proceduralisation processes that lead to fluency. In order to extensively practise the language however, learners should have increased willing to communicate; this construct is discussed next.

2.5. Willingness to Communicate

Willingness to communicate (WTC) is a construct which accounts for the probability of someone to speak 'when free to do so' (MacIntyre, 2007: 564). When people are faced with a situation where they can communicate, some choose to speak while others choose to remain silent. This choice is what WTC is concerned with. The construct was originally introduced by McCroskey and associates (McCroskey & Baer, 1985; McCroskey & Richmond, 1987; McCroskey, 1992) in the communication literature and it concerned L1 communication. According to them, WTC was a stable personality trait, as it tended to be the same in various situations. The model took into account the effects of certain variables, such as apprehension, reticence, introversion and shyness on the communicative behaviour of speakers (Yashima, 2002; Yashima *et al.*, 2004).

2.5.1. MacIntyre's models

It has been suggested in research that L2 WTC is not simply a manifestation of L1 WTC (e.g. Charos, 1994), mainly because the communicative competence in the L2 context may vary from 0% to 100% (MacIntyre, Clément, Dörnyei & Noels, 1998). Learners are often required to communicate based on their 'underdeveloped L2 skills' (MacIntyre & Legatto, 2011: 149). This means that other variables (e.g. anxiety, confidence, perceived communication ability)

which do not emerge in L1 communication affect the learners' choice of whether to speak or not.

In order to create a WTC model for L2 communication, MacIntyre (1994) adapted McCroskey (1985 in MacIntyre *et al.,* 1998) model by proposing a path model. According to MacIntyre (1994), L2 WTC had two key predictors: perceived communicative competence and communication anxiety. In other words, when L2 learners have high perceived communication competence and low communicative anxiety, then their WTC is higher. Four years later however, MacIntyre, Clement, Dörnyei and Noel (1998) re-conceptualised WTC from a more theoretical perspective. This model was a pyramid-shaped model which included a wider range of variables that could influence WTC (see Figure 2.5.1 below).

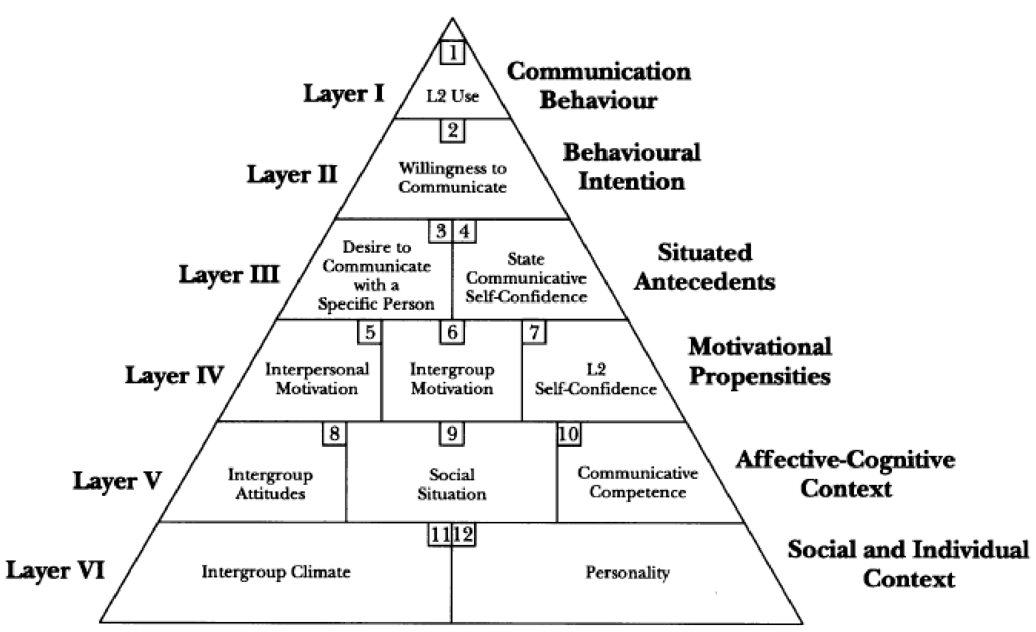


Figure 2.5.1. MacIntyre *et al.* (1998) pyramid model of WTC (p. 547)

This model represents the influences that affect one's choice when faced with the decision of speaking or not. The bottom three layers represent *stable, enduring influences,* such as the personality of the speaker or intergroup climate. These are placed at the bottom of the pyramid, as they form the foundation on which other kinds of influences operate (MacIntyre

et al., 1998). The next level of influences includes motivational propensities (e.g. motivation, self-confidence), affective-cognitive influences (e.g. intergroup attitudes, social situation, communicative competence) and social or individual influences (e.g. personality and intergroup climate). The top three layers represent *situation-specific influences*, such as desire to speak to the interlocutor or knowledge of topic. These include situation-specific influences (e.g. desire to communicate with interlocutor, state communicative self-confidence), behavioural intention, which is WTC or in other words, the state of readiness of the speaker to enter into discourse at that time. Finally, the top layer considers the communication behaviour of the speaker in the L2. This includes activities like speaking in the class, reading L2 newspapers, watching L2 television and so on.

As MacIntyre (2007) puts it, the latest model of WTC suggests that 'the initiation of communication is a matter of choice, a decision to be made at a particular moment' (569). In the latest model therefore, WTC is conceptualised as a state of readiness or 'an act of volition' (MacIntyre, 2007: 569). The learner makes a conscious choice of whether s/he is willing to communicate or not, depending on the conditions summarised in the pyramid model. So in the present study, WTC is not only regarded as a choice of the learner to initiate communication, but also the extent to which the learner is willing to sustain this communication.

2.5.2. Measuring WTC

The concept of WTC has been measured with self-reports in the past. A number of fairly recent studies adapted McCroskey's (1992) scale of L1 WTC to measure this L2 WTC. This scale asked participants to rate from a scale of 0% to 100% the percentage of time they would choose to communicate in four different contexts (public speaking, talking in meetings, group discussions and interpersonal conversations), with three different types of interlocutors

(stranger, acquaintance and friends). Twelve items from this measure were borrowed by Ghonsooly, Khajavy and Asadpour (2012), who wanted to examine the L2 WTC of 158 non-English major university students in Iran. McCroskey's (1992) scale was used by other researchers within the EFL context (e.g. Yashima, 2002; Hashimoto, 2002).

Yashima, Zenuk-Nishide and Shimizu (2004) also used McCroskey's (1992) scale in two investigations with Japanese high school students. In the first investigation, 160 students participated in a school programme, which involved being taught by two teachers; one of which was a native English speaker. The students completed a questionnaire measuring attitudes, motivation and WTC. As in the previous study, WTC was measured with McCroskey's (1992) adapted scale. This asked students to indicate the percentage of time they would choose to communicate in twelve different situations (four types of contexts and three types of interlocutors). Other measures in the questionnaire included frequency of communication, perceived communication competence and communication anxiety. Yashima *et al.* (2004) found that higher WTC scores strongly correlated with communicating more in the classroom, asking their teachers more questions outside the classroom and having higher perceived communication competence.

In their second investigation, fifty-seven students, who participated in a study-abroad programme in the US, were recruited. These students completed two different questionnaires. The first questionnaire, which was administered before their departure, contained the same measures as the questionnaire in the first investigation, that is McCroskey's (1992) adapted scale for WTC. The second questionnaire, administered three weeks after their arrival in the US, included questions on the amount and frequency of their communication with their host families, who were native English speakers. The scale assessing frequency of communication was a 5-point scale from *not at all* to *very frequently* and it included statements such as "I

reported to my host family what happened at school" (Yashima *et al.,* 2004: 137). The results confirmed the ones of the first investigation. In addition, it was found that WTC before departure significantly correlated with the frequency and length of communication with host families items.

In a study looking at the potential effect of views on language learning and social support on WTC, MacIntyre, Baker, Clément and Conrod (2001) used a questionnaire to measure WTC, that was not based on McCroskey's (1992) scale. The WTC of seventy-nine high schools students of L2 French immersion living in an English-speaking community in Canada was measured. The administered questionnaire had four parts. The first part measured WTC inside the classroom by asking student to rate on a scale from 1 to 5 how willing they are when engaging in a range of communication tasks. The second part of the questionnaire measured WTC outside the classroom, using the same scale. The third section measured orientations for language learning and the final part measured social support. A positive correlation was found between orientations for language learning and WTC, both inside and outside the classroom. Also, social support played an important role for levels of WTC outside the classroom.

In a recent study, Freiermuth and Huang (2012) examined the motivation of twenty Japanese university students of English who were asked to solve tasks, through an online chat software with nineteen Taiwanese university students of English. In this study, WTC was considered to be one of the factors affecting motivation, along with three other variables including task attractiveness, task innovation and the need to use the target language. In order to examine whether the given task had an effect on these four variables, the students completed a questionnaire after the session, which included questions about the effect of the four variables on motivation.

Having analysed the questionnaires qualitatively, the researchers found references of the students to three of the variables included in MacIntyre *et al.'*s (1998) model of WTC: confidence, anxiety and power differences. Considering confidence first, the students' responses showed that the online chat helped them feel confident while communicating in English. As for anxiety, online chatting made some students feel less nervous than face-to-face communication, and others more nervous. Finally, in terms of power differences, the Japanese students found that the Taiwanese students were a bit aggressive, while the Taiwanese students thought that their Japanese peers were more proficient. However, according to the researchers, these two factors, not only did not hinder communication, but were rather perceived as motivating factors. The researchers concluded therefore that WTC was not affected by perceived status differences.

In a study in Canada, MacIntyre and Doucette (2010) measured the WTC of 238 English-speaking high school students (14-18-year-olds), who were studying French. The aim of their study was to examine the extent to which individual differences between learners in terms of their action control (hesitation, preoccupation and volatility) predicted WTC. The students completed a questionnaire which included measures on all three variables of action control, as well as measures on communication; part of which was WTC measures. In terms of trait-like WTC (McCroskey & Baer, 1985), students were asked to rate from a scale of 0% to 100%, the average percentage of time they would be willing to communicate in French and English in different situations. In terms of WTC inside and outside the classroom (MacIntyre *et al.,* 2001), students were asked to give a percentage of time they would be willing to communicate in the L2 both inside and outside the classroom. The results supported the authors' prediction on the relationship between action control and WTC, so they argued that 'the theory of action control represents a new avenue of investigation into WTC research' (MacIntyre & Doucette, 2010: 168).

However, a year later MacIntyre and Legatto (2011) argued that the methods for measuring WTC to date have not been able to capture the dynamic nature of WTC. They proposed a new methodology therefore, which was still based on self-reports, but it was designed to record the constant changes of WTC. Six young adult female English-speakers, who were university students learning French, tested this new methodology. Their proposed methodology consisted of two main parts. The first part measured the stable influences of WTC through a questionnaire, which 'assessed the respondents' usual willingness to initiate communication' (MacIntyre & Legatto, 2011: 154), including questions on extraversion and language anxiety. The second part measured situation-specific influences by asking participants to complete eight communicative tasks individually in the presence of a research assistant while being video-recorded. Immediately after task completion, the recordings were played back to the participants and, with the help of the research assistant and a software, the participants completed a moment-by-moment rating of their own WTC. The software then produced a graph that showed fluctuations in WTC. Looking at the graphs, the reasons these changes in WTC took place were discussed with the participants.

It is also worth mentioning that MacIntyre and Legatto (2011) measured speaking time as well by counting the seconds each participant spent on each task. However, they found a near zero correlation between speaking time and dynamic WTC ratings over the eight tasks. Nevertheless, it should be noted that WTC ratings were based on self-reports; in other words, the perceptions of the participants on their level of WTC. The problem with self-reports is that people differ in the way they perceive things; what is high WTC for one person may be an average WTC for another person. Self-reports therefore cannot be a reliable WTC measure.

2.5.3 Conclusions on WTC

Researchers seem to agree on MacIntyre *et al.*'s (1998) pyramid model of WTC in terms of L2 communication, as this model captures both stable and situational influences. However, although it is clear what the concept of WTC is and what it represents, measuring it has not been an easy task due to its dynamic nature. The most widely used measure of WTC is self-reports; in other words, participants were asked to report in some way how willing they were to speak. The majority of studies used McCroskey's (1992) WTC scale (e.g. Yashima, *et al.*, 2004; Ghonsooly *et al.,* 2012), which asked participants to give the percentage of time they would be willing to speak in twelve different situations (different contexts and interlocutors). MacIntyre and Legatto (2011) however, rejected this method as it was unable to capture the constant changes of WTC. Basing their methodology on self-reports again, they proposed reporting moment-by-moment changes in WTC as a more accurate measure.

However, the problem with all the above studies is that they all use self-reports. What self-reports actually measure is the perceptions of the participants on their WTC, rather than WTC itself. Perceptions cannot be a standardised unbiased measure, as what is one person's high WTC could be another person's average WTC. In terms of reporting percentages of time, one might think that s/he speaks 80% of the time in the L2, when in fact they speak 60%.

Nevertheless, it is worth noting that McCroskey's (1992) scale, used in the majority of these studies, is related to *length of speaking time*. As discussed earlier, MacIntyre and Legatto (2011) was the only study who attempted to examine speaking time as a WTC measure, but found a negative correlation between speaking time and WTC ratings. However, as argued above, WTC ratings should not have been used as a solid WTC measure, as once again they were based on self-reports. It is argued therefore that speaking time alone would be a more

reliable measure of WTC, as it captures how much a speaker is willing to expand in a certain situation, and this is one of the ways WTC is measured in the current study.

2.6. Hypotheses

Based on the discussion and the arguments that have been put forward in the review of the literature, the rationale behind the hypotheses of the present study is discussed here. Having established that WTC is a matter of choice depending on the context, it could be hypothesised that by allowing students to codeswitch, WTC will be enhanced. With increased WTC, learners will be more likely to choose to communicate in the L2, if they are given the choice to switch codes in case of difficulty. This is also an idea that was put forward by Macaro *et al.* (2012) in the paper describing the study with bilingual assistants. The authors suggested that 'being allowed to code-switch when speaking might bring about a greater willingness to communicate' (Macaro *et al.,* 2012: 4). It is further predicted here that learners will even attempt to express more complex ideas, knowing that they can rely on their L1. Finally, allowing codeswitching will enable students to maintain the flow of their conversation and to prevent any communication breakdowns.

Moreover, codeswitching will provide cognitive support during task completion, as it allows for metalanguage to take place. As discussed in the task-based learning section, metalanguage enables higher cognitive activities, such as planning, organising and monitoring. As it has been argued in this review, a condition that simply permits codeswitching could provide cognitive support during task completion, but it could also result in L1 overuse. The possibility of students abusing this freedom of codeswitching could result in the task being completed in the L1 with L2 switches for task language. This undesirable result comes into conflict with the aim of tasks, which is effective L2 communication.

In order to overcome this problem, the *gradual recycling process* of the L1 use into the L2 is proposed. This is expected to move learners to a state of less reliance on the L1. No theory to date suggests that L2 metalanguage cannot offer the same cognitive support as L1 metalanguage. In fact, the study conducted by DiCamilla and Antón (2012) shows that the same functions that are served by the L1 of beginner learners are expressed in the L2 by advanced learners. This suggests that as the proficiency level of learners increases, the L2 is used for the same functions that the L1 was previously used for. Based on this argument, it seems logical to suggest that by recycling the use of the L1 into the L2, this desirable result can be achieved much faster and perhaps much more efficiently.

It is further predicted that with increased WTC and L2 metalanguage, task procedures will be facilitated and L2 practice will be increased. According to Anderson's ACT model (1983), this extensive practice facilitates proceduralisation processes which, as explained above, are crucial for oral fluency development. It is hypothesised, thus, that by allowing students to codeswitch, L2 oral fluency will eventually develop. All these processes predicted in this hypothesis are represented graphically in figure 2.6.1 below.

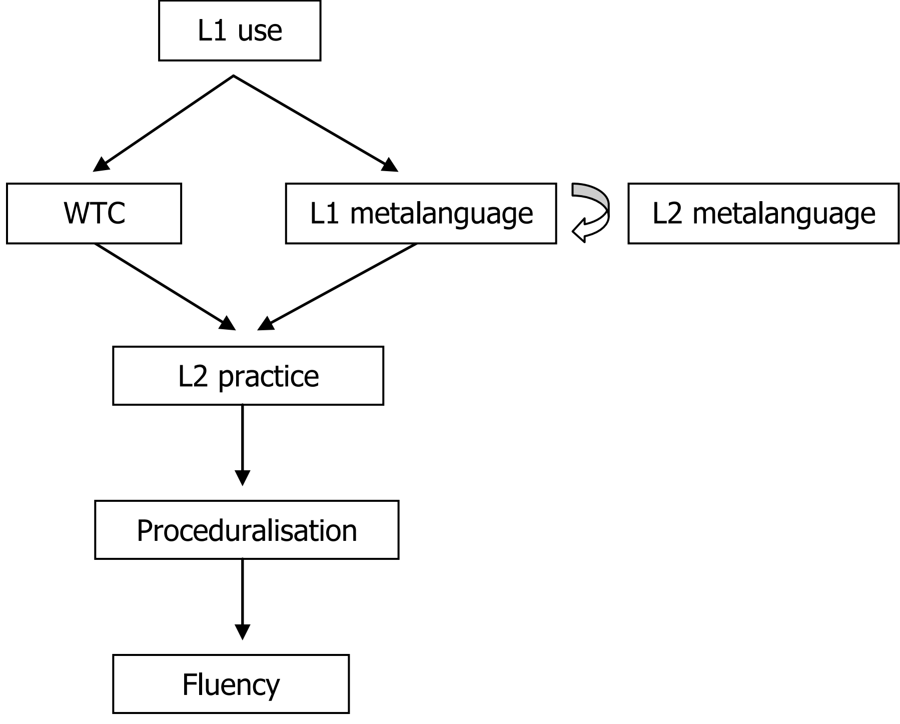


Figure 2.6.1 Graphic representation of hypothesis

From the discussions in this review, it seems that variables in language acquisition are strongly linked and therefore manipulating one variable could also affect other variables. An example of this link emerges from the hypotheses of the present study which predicts that by allowing codeswitching, WTC will be enhanced and metalanguage will occur. With metalanguage recycled into the L2, these two variables will lead to extensive L2 oral practice, which will then facilitate proceduralisation processes that will lead to fluency. This linear relationship between the variables of this hypothesis provides evidence that progress in language learning should be studied as a blend of variables which constantly interact, rather than as isolated variables.

2.7. Chapter summary

This chapter presented a review of research carried out in four areas: task-based learning, classroom codeswitching, oral fluency and willingness to communicate.

From the discussion on task-based learning, it emerged that, while this was a very influential approach in language teaching methodology, there were a number of problems, both in research and in its classroom implementation that needed to be resolved. Problems in research concerned the lack of consensus on the definition of tasks, while problems reported by teachers regarding task implementation included lack of time, L1 use and lack of discipline. A significant part of the discussion focused on task repetition, which is an important feature in the present study. It was concluded that, despite the methodological problems of task repetition studies, there was some evidence suggesting that task repetition positively affected fluency and complexity, but not accuracy. The task repetition plus feedback package was proposed in the present study, as a more viable way to use repetition in real classroom contexts.

The second part of this review was dedicated to classroom codeswitching. Although this was an issue of controversy for years, it has been studied extensively in more recent years. Empirical studies focused on teacher codeswitching and proved that, regardless of the context, use of the L1 existed to some extent and served a range of social and pedagogical functions. Studies on the perceptions of teachers and students on classroom codeswitching confirmed this result by acknowledging and accepting the presence of L1 in their classrooms. Experimental studies on teacher codeswitching provided some evidence suggesting a positive relationship between codeswitching and vocabulary acquisition. This section, however, focused on learner codeswitching, which was analysed mainly within Vygotsky's (1985) sociocultural theory framework in the past, and was proven to serve various social, cognitive and pedagogical functions. While researchers argued that for these reasons, L1 use during L2

interactions should be allowed, but not necessarily encouraged, the present study attempted to take this a step further by suggesting that something should be done to make learners rely less on their L1. This is further discussed as part of the hypothesis.

From the third part of the review, which focused on fluency, it emerged that the complicated nature of fluency had resulted in a variety of proposed fluency measurements over the years. Speech rate and mean length of runs, however, were considered to be the best predictors of fluency. As for fluency development, Levelt's (1989) model of speech production and Anderson's (1983) Adaptive Control of Thought model were used to explain how speech becomes fluent.

The final research area this review was concerned with the concept of WTC. While researchers seemed to agree on what WTC actually was, the controversy seemed to revolve around the way in which the concept could be measured. Researchers used various types of self-report measures in the past which, as argued in the present review, may not have been able to measure WTC itself, but rather the perceptions of speakers on their WTC. As most of these self-report measures included a measurement of the perceived length of speaking time, the actual speaking time was proposed as a better measure of WTC.

The discussion of these four research areas led to the formation of the hypothesis for the present study. In short, the hypothesis suggested that allowing codeswitching would lead to the increase of WTC and L1 metalanguage. With the gradual recycling of L1 metalanguage into the L2, L2 oral practice would increase and proceduralisation processes required for fluent speech would be facilitated.

CHAPTER 3: Methodology

3.1. Introduction

Following the review of literature and the hypotheses of the present study, this chapter starts by presenting the research questions. A discussion on the appropriate research design follows. The chapter continues with a description of the population and the sampling procedure. Then a detailed record of the research study and the data collection process follows. The results of a pilot study are presented in support of the research design. In addition, the data analysis procedure followed for each research question is described. The chapter ends by referring to the limitations of the design.

3.2. Research Questions

Following the hypotheses that emerged from the discussion of the literature, five research questions were formulated. The first research question aimed to introduce the context of the study in more depth, with particular focus on task-based learning. As there is evidence suggesting that task implementation has been an issue in some countries (e.g. Carless, 2004), this question aimed to investigate the opinions of teachers of English in Cyprus regarding this matter. In addition, their opinion on the "controversial" matter of task repetition was also sought. Another important issue that emerged from the discussion of the literature was that of the role of metalanguage and the language it occurs in. Investigating the teachers' opinions on this matter was judged to be crucial. Finally, this research question inquired about the role of proficiency pairing as, based on Vygotsky's SCT theory, it emerged that more proficient students may offer scaffolded help to their less proficient peers. It would be interesting to know the opinion of teachers therefore on how students are usually paired in group work. In general, this research question was also added in order to ascertain whether a pedagogy

involving tasks can feasibly be introduced or developed in this context. The research question was formulated as follows:

1) What are teachers' views on task implementation in Cyprus, task repetition, metalanguage and the role of proficiency pairs during task completion?

Apart from setting the context of how task-based learning is approached in Cyprus, this research question was expected to provide a foundation, on which the quasi-experimental part of the study could be built. In other words, a more profound understanding of the context was considered to be essential prior to the construction of an intervention. In addition, approaching the issue from both perspectives, that is the interviews and the intervention, can offer some kind of validation to the results when these are discussed later on.

One of the main issues emerging from the literature of task-based learning concerned the effect that task repetition could have on learners' output. According to Bygate (2001), when speakers attempt a task for a second time, their attention is freer to focus on other features of speech, such as fluency, accuracy or complexity, because the formulations had been previously processed and stored. Bygate (2001) argued that the minimum that can happen with task repetition is the improvement of fluency. Based on this argument, the second research question aimed to examine whether the content of the output of learners was improved in any way when repeating the same task. The question was the following:

2) How does the output change from session 1 (S1) to session 2 (S2) particularly in terms of content?

The following research question investigates the task repetition plus feedback package (hereafter, TR+) proposed in the literature review as a more pedagogically justified approach. Two types of this package are tested: the first involves repetition under CS conditions with feedback coming in the form of recycling the L1 metalanguage into the L2, as proposed in

the hypothesis; and the second involves task repetition under English-only conditions with feedback on accuracy. The research question investigates the effects of these packages on willingness to communicate (WTC), as it was predicted that the first package involving CS would increase WTC. The question however is divided into two sub-questions: the first investigates whether TR+ leads to more WTC than no task repetition; and the second investigates whether there is a difference in the effect of the two different types of TR+ on WTC. The research question was formed as follows:

3) a) Does task repetition plus feedback lead to more willingness to communicate than no task repetition?
   b) Does task repetition with codeswitching *plus* lead to more willingness to communicate than task repetition with English-only *plus*?

The next research question examines whether TR+ has an effect on oral fluency. As it was hypothesised, a potential increase of WTC, investigated in the previous research question, in combination with recycled L2 metalanguage, would lead to extensive L2 practice, which would facilitate proceduralisation processes, required for fluency development. Again, this question was divided into two parts. As in the previous research question, first it investigated whether TR+ leads to more fluency than no task repetition. Then, it further inquired about the effect of the two types of the TR+ package on fluency. Nevertheless, the question assumed that a potential fluency development would come with no effects on accuracy.

4) a) Does task repetition plus feedback lead to more fluency than no task repetition?
   b) Does task repetition with codeswitching *plus* lead to more fluency than task repetition with English-only *plus*, with no detrimental effects on accuracy?

Finally, as the results of the pilot study will show later on in this chapter, student proficiency seems to vary greatly in this context. It would be interesting to examine therefore, whether

TR+ is a package that benefits high proficiency students more than lower proficiency students. The final research question addressed this issue:

5) Do higher proficiency students benefit more from task repetition plus feedback than lower proficiency students?

3.3. Research design

For the present study, a mixed methods research design has been adopted, in order to combine both quantitative and qualitative methods. Increasingly seen as the third approach of research methodology (Dörnyei, 2007), this combined approach was introduced in the field of social sciences in the 1970s, with the concept of *triangulation;* the combination of different sources of data studying the same phenomenon (Dörnyei, 2007).

The combination of the two approaches is more likely to provide a better understanding of the phenomenon under investigation. This is mainly because mixed methods overcome the weaknesses of each approach, when that is used on its own. More specifically, quantitative approaches were criticised for being simplistic, de-contextualised and unable to capture the true meaning of the data; while qualitative approaches were criticised for being very context-specific and for using samples that are not representative enough (Dörnyei, 2007). Using both approaches compensates for these weaknesses and contributes to increasing the internal and the external validity of the study (Dörnyei, 2007).

However, the mixed methods approach has not been spared of criticism. Dörnyei (2007) mentioned that mixed methods may be used as a replacement for conceptual and insightful thinking, when a researcher is in doubt about the results of a single approach. Moreover, Hesse-Biber and Leavy (2006) referred to the issue of how well-trained is the majority of researchers in conducting research in both paradigms. The fact is that most researchers

specialise in either one or the other, and this should not be disregarded. Despite these criticisms, mixed method approaches have been increasingly used during the past two decades because of the insights they can provide to the study of a phenomenon. This approach therefore was used in the present study as well.

Concerning the quantitative aspect of the study, the most ideal research design is an *experimental design,* which can establish cause-effect relationships. A widely used experimental design is the *Randomised Control Trial (RCT)* or the *intervention study*. This involves at least two groups: an *experimental* or *treatment group,* which receives some kind of manipulation or treatment by the researcher, and a *control* or *baseline* group, which follows its normal practice and it is used for comparison purposes. The dependent variable is measured with *pre-tests*, administered before the intervention, and *post-tests*, administered after the intervention (Dörnyei, 2007). The comparison of the results of these two tests establishes whether the experimental manipulation has an effect. A *delayed post-test* is sometimes employed in order to follow the potential effect of the intervention over a period of time.

Nevertheless, firm cause-effect relationships are very hard, if not impossible, to establish in the social sciences because of confounding variables which may interfere in the situation. For this reason, the *random assignment* of the participants to groups is essential in this kind of designs (Dörnyei, 2007). The existence of a control group and randomisation are the two characteristics of pure experimentation, which ensure that the manipulated independent variable has the sole responsibility for the variation of the dependent variable (Bryman, 2008). However, true experiments are not always feasible in educational contexts. Nevertheless, their consideration is crucial as they are 'often used as a yardstick against which non-experimental research is assessed' (Bryman, 2008: 35).

As the ultimate aim of the present study was to establish a relationship between learner codeswitching and L2 oral fluency, a pure experimental design study would have been the ideal design. However, for practical reasons, this was not possible mainly because parents and students were not willing to participate in a study that would take up time outside of the school. Where true experiments are not possible, *quasi-experimental designs* are used; as was done in the present study. These are similar to true experiments in all aspects, except for randomisation, which is replaced with *intact classes* in educational contexts. Although this undermines the internal validity of the design, it strengthens the ecological validity of the study, as there is no laboratory environment.

There are ways of strengthening quasi-experimental designs: first, by preventing participants from choosing the conditions they want to be in; and second, by making an effort to minimize the pre-test differences between the groups prior to the intervention (Heinsman & Shadish, 1996). The latter can be achieved by *matching the participants* of the groups on a case-by-case basis. Although perfect matching is almost impossible, the groups become more comparable (Dornyei, 2007). Also, pre-test differences can be statistically screened with an *analysis of covariance* (ANCOVA). Both these ways of strengthening the design were used in the present study.

It is worth clarifying at this point the independent variables and the dependent variables in the study. Concerning the independent variables first, research questions 3 and 4 inquire about the effects of three conditions. These are: TR+*;* task repetition with CS *plus*; and task repetition with EO *plus.* These conditions therefore become the three independent variables. Looking at the dependent variables, the hypothesis predicts an effect on willingness to communicate, which will consequently lead to extensive oral practice and therefore to the required proceduralisation processes for fluency development. Research question 4 however, assumes

that fluency development would come with no detrimental effects on accuracy. The hypothesis also predicts that, with the support of L1, learners will attempt to express more complicated ideas. Therefore, there are four dependent variables in the study: willingness to communicate, content, fluency and accuracy. The table below shows these variables more clearly.

Table 3.3.1 Independent and dependent variables

Independent variables	Dependent variables
Task repetition *plus* Task repetition with CS *plus* Task repetition with EO *plus*	Willingness to communicate Content Fluency Accuracy

In terms of the qualitative aspect of the research design, interviews and recordings of students' interactions were used as methods to collect qualitative data. These data were used to respond to the first and second research questions.

3.4. Population - Sample

The target population of the study was sixth grade students of Greek-medium state primary schools in urban Nicosia. Young learners were chosen for two reasons: first, research on the codeswitching behaviour of non-adult learners is more limited; and second, access to primary schools was a more convenient option for me. In order to recruit schools that fit these criteria, a *non-probability convenience sampling* procedure was adopted, according to which, the sample available to the researcher is selected. Although this sampling method may

compromise the generalisability of the results, these studies could lay a foundation for further research or allow links to be made with the findings of previous research (Bryman, 2008).

Three Greek-medium state primary schools, which I had previous contact with, had been contacted. These were all urban schools, situated close to each other in Nicosia. Two of the schools belonged to the same area and shared the same English teacher. For this reason, it was decided that these two schools would act as the two experimental groups. The decision on which school would act as each experimental group was taken randomly (e.g. flipped a coin). The third school belonged to a different catchment area (although not very distanced from the other two schools) and acted as a comparison group. This should be taken into consideration when interpreting the results.

In terms of school performance, the Ministry of Education and Culture of Cyprus does not prepare any school performance tables for state schools. All headteachers and teachers are employed by the government and they are randomly allocated to schools. The three schools therefore cannot be compared in terms of performance.

Sixth grade students were chosen as the target population. As the number of students in classrooms is restricted by the Ministry to twenty-five, the first school had two classrooms of 19 students each; the second had three classrooms of 20 students each; and the third had two classrooms of 25 students each. All the classes were asked to participate in the study. From a total of 148 potential participants however, only half of the students agreed to participate. Some students did not wish to participate because they did not feel confident with their English, while the parents of others did not agree to give their consent.

Upon my request, the headteachers of the schools made another effort to persuade or remind the students to return the consent forms completed. The comparison group, which was the school with the least participants, was informed that a small donation would be made to the

school at the end of the study, if more students would participate. Although this did not increase the number of participants significantly, a small donation of English books was made to the library of the school. Eventually, a total of 75 students from all three schools agreed to participate in the study.

Table 3.4.1 Number of participants

Codeswitching group	English-only group	Comparison group	Total
21	35	19	**75**

As shown in Table 3.4.1 above, the number of participants in each group was uneven. The reason is because the school of the second experimental group was a larger school, so it was more likely to recruit more students from that school.

Overall, the sample matches the population very closely. All the participants recruited were sixth grade student who attended Greek-medium state primary schools in urban Nicosia.

3.5. The intervention

The intervention contained three groups: two experimental groups and a comparison group[1]. Each condition was assigned to a different school for two reasons: first, this would increase the possibilities of having a larger sample; and second, the confounding variable of the contamination of the conditions between students would be eliminated. All sixth grade classes of these three schools participated in the study and remained intact. The intervention lasted eight weeks during the normal classroom time of the EFL lessons; that is two lessons per

[1] The label *comparison group* is preferred in such designs, as the absence of randomisation suggests that extraneous variables might not be controlled for. The term *comparison* suggests that the performance of this group will be compared to the others.

week for each class. During the intervention, the English teachers of these classes taught the lesson as normal by using their usual book.

For the purposes of this study, the students of all three groups completed the same tasks, in dyads or small groups of three[2], during the last ten minutes of each lesson. A different task was completed every week. All students in these classes, including the ones who were not participating in the study, completed the tasks as part of their lesson. However, only the ones with consent were audio-recorded. With evidence suggesting that interlocutor familiarity affects performance (e.g. O'Sullivan, 2002), students were asked to choose their interlocutor from the students that had given consent, and to remain with the same interlocutors until the end of the intervention.

What differed between the three groups were the conditions under which each group completed the tasks; these are described in section 3.5.2 below. An oral production test (see section 3.5.3) was administered as a pre-test, post-test and delayed post-test, in order to measure the L2 oral fluency of the participants.

3.5.1. Tasks

The tasks used in the study were designed by me based on the book that the students were using at the time of the study. The book, called *English for Communication 3* (Georgiou, Kitromilides & Englezaki, 2007), was written and published by the Ministry of Education in Cyprus. It was quite an old book; published in 1994 for the first time and then revised three times after that (in 2001, 2006, 2007). The pupil's book was accompanied by a workbook and a teacher's book. As stated in the teacher's book, the aim of the book was to 'enable learners

[2] Groups of three were only formed when the number of participants in a class was odd.

to communicate effectively in various situations' (Georgiou, Kitromilides & Englezaki, 2007: 6). The book consisted of twenty-five units, each introducing a new grammatical structure through a short text. The texts were then followed by several activities; either listening activities, 'work in pairs' activities, games or writing activities, which aimed to practise the grammatical structure at hand.

As each unit from the book requires approximately both lessons of a week to be covered, I designed one task for each week. When designing the tasks, my aim was to remain within the boundaries of the structure introduced in each unit, while making the task as open-ended as possible to allow creativity. The reason I wished to relate my tasks to the curriculum was firstly because this would ensure that the students had the required knowledge to complete them, and secondly because the school, the teachers and the parents would be more content knowing that the students did not deviate from the curriculum.

A grammatical structure that was emphasised in these books, with four chapters dedicated to it, was the past tense. The tasks designed for the first four weeks therefore, aimed to induce learners to narrate past events. The first two tasks were a narrative task (Appendix 1) and an information gap task (Appendix 2), which asked learners to have a short conversation about what they did the previous day based on the pictures. For the following two weeks (Appendices 3 & 4), the students were asked to narrate the story presented by two different sequences of pictures in the past tense. Some verbs or nouns were given as cues. As the students completed these in pairs, they were asked to narrate each picture alternatively so that they would both contribute to the final outcome.

For the next four weeks of the intervention, the structures introduced in the chapters varied. In week 5, students built on their previous basic knowledge of the comparative and superlative forms. An opinion gap task (Appendix 5) asked learners to discuss which would be the best

way for them to travel from London to Paris by using the pictures of several means of transport and the adjectives given. For weeks 6 and 7, students had to implement the “should/shouldn’t” structure in two role-plays: a doctor/patient roleplay (Appendix 6) and a dentist/patient roleplay (Appendix 7). With the help of given pictures, the pairs had to engage in a conversation with the doctor or dentist giving advice to the patient and the patient asking questions about what he/she should or shouldn’t do.

For the final week, students were introduced to the imperative form, by practising giving and following instructions. In an information gap task (Appendix 8), student A and student B were placed on different spots on the same map. The purpose of the task was to find each other by giving instructions.

3.5.2. Conditions in the experiment

While all three groups completed exactly the same tasks, they differed in terms of the conditions under which they completed them. As already mentioned, the experiment consisted of two experimental groups and one comparison group. The two experimental groups tested two different pedagogical packages, designed based on the hypotheses of the present study.

i. <u>Codeswitching group (Experimental)</u>

The first experimental group, the *Codeswitching group (CS group)*, tested pedagogical package A, which had three components:

Codeswitching + Feedback on metalanguage + Task repetition

This group was given a task to complete in pairs during the last ten minutes of session 1 (S1) of each week. Starting with the *codeswitching* component, this package offered them the

opportunity to switch to the L1 whenever this was necessary, while completing tasks. This L1 use would allow metalanguage to occur, providing additional cognitive support to solving tasks. This suggests that students in this group would be in a better place to attempt expressing more complex ideas. Moreover, since WTC is more a matter of choice depending on the context, by allowing students to codeswitch, it was expected that WTC would be enhanced in this group. With increased WTC, students would be more likely to choose to communicate in the L2.

As suggested in the hypothesis of the present study, the *feedback on metalanguge* component of this pedagogical package attempts to recycle the L1 metalanguage into the L2. Having recorded the students' interactions in S1 of each week, the most frequently occurring L1 metalanguage words and phrases were collected and translated into the L2 by me. A handout with this translated language was prepared and distributed to the students of the whole classroom (about 20 students since the ones who would not be recorded also participated) during the last ten minutes of S2 (see example in Appendix 9). As the time was limited, I asked them to quietly read through their handout once and then I read out the new L2 phrases to them. If more time was available, the students were asked to raise their hands if they wanted to read out the new phrases themselves. This was believed to help them engage with the feedback more easily. Having gone through the handout, the students repeated the same task, but this time they were encouraged to use the newly introduced language. The same procedure was followed for eight weeks.

With increased WTC and with L2 metalanguage, task procedures were expected to be facilitated and L2 practice would increase. Based on Anderson's (1985) ACT model, this extensive L2 practice facilitates proceduralisation processes, which are crucial for oral

fluency development. Therefore, this group tests the hypothesis predicting that allowing students to codeswitch may lead to L2 fluency development.

ii. English-only group (Experimental)

As the fluency of the CS group was expected to develop, this should come with no detrimental effects on accuracy. The main purpose of the second experimental group was to ensure that by allowing codeswitching accuracy does not deteriorate; or it does not deteriorate more than it would if these students were in a non-codeswitching condition. The students of the second experimental group, the *English-only group (EO group)*, tested a different pedagogical package, which also had three components:

English only + *Feedback on accuracy* + *Task repetition*

The students in the EO group were given the same tasks as the CS group to complete during the last ten minutes of S1 of each week. However, this group was asked to complete them in English only. The English-only component was expected to prevent metalanguage, restrict WTC and possibly hinder L2 output, as discussed in the literature review (e.g. Macaro *et al.,* 2012).

In order to match the *plus* component of the CS condition, this group was given feedback on accuracy. Once the interactions of S1 were recorded, a selection of the most common errors was made and a handout was prepared with their corrected versions (see example Appendix 10). In other words, the students' productions, which contained some kind of error, were reformulated and presented to the students error-free at the beginning of the last ten minutes of S2. As with the CS group, all students in the EO classrooms were asked to read the handout once quietly. Due to lack of time, I then read out the corrected versions of the utterances by pointing out where the errors were. In case we had more time available, I asked the students

instead to point out the errors by raising their hands to give the answer. After going through the handout, the students repeated the same task, but they were encouraged to use the language of the handout. Again, this procedure was followed for eight weeks.

iii. Comparison group

Finally, the students of the third group, the *Comparison group (COM group)*, were asked to complete each task once, without being given any specific language instructions or any kind of feedback. This group was added in order to be used as yardstick against which TR+ would be compared. Table 3.5.1 below shows a graphic representation of the research design.

Table 3.5.1 Graphic representation of design

		CONDITIONS		
		Codeswitching group (School 1)	**English-only group (School 2)**	**Comparison group (School 3)**
WEEK 0		Pre-test		
WEEK 1	**Session 1**	Task 1 (CS)	Task 1 (EO)	Task 1 (No instructions)
	Session 2	Task 1 (Recycling on fluency)	Task 1 (Recycling on accuracy)	---------
WEEK 2	**Session 1**	Task 2 (CS)	Task 2 (EO)	Task 2 (No instructions)
	Session 2	Task 2 (Recycling on fluency)	Task 2 (Recycling on accuracy)	---------
WEEK 3	**Session 1**	Task 3 (CS)	Task 3 (EO)	Task 3 (No instructions)
	Session 2	Task 3 (Recycling on fluency)	Task 3 (Recycling on accuracy)	---------
....	**...**	...	...	...
WEEK 8	**Session 1**	Task 8 (CS)	Task 8 (EO)	Task 8 (No instructions)
	Session 2	Task 8 (Recycling on fluency)	Task 8 (Recycling on accuracy)	---------
WEEK 9		Post-test 1		
WEEK 10		Easter break		
WEEK 11				
WEEK 12		Post-test 2		

3.5.3. Oral production tests

All participants took an oral production test individually at three time points during the study. The first test, the *pre-test,* was taken before the intervention. The aim of this test was twofold: first, to determine the level of fluency between the groups prior to the intervention; and second, to show any potential effects of the intervention, when compared to post-test results. The second oral production test, *the post-test,* was taken immediately after the intervention. This test would be able to detect any immediate effects that the intervention had on any of the three groups. After the first post-test, the Easter break lasted for two weeks, during which none of the students could have had any exposure to formal English instruction. The third test, *the delayed post-test,* took place at the end of these two weeks, when the students returned to schools. The aim of this test was to show whether any potential gains from the intervention would be retained two weeks later.

Two types of speaking tests have been used in the past: *direct* and *indirect tests* (Fulcher, 2003). In *direct* testing, test takers are involved in human interaction. The first published test of speaking was a direct test, namely the *Foreign Service Institute's (FSI) Oral Proficiency Interview (OPI)* (Fulcher, 2003). The OPIs were conducted by a native-speaker interviewer, while a rater was observing and grading the test taker. The interview lasted between ten to forty minutes. It began with 'simple social formulae' (Fulcher, 2003: 11), it continued with eliciting grammatical structures and it finished with a role-play or a dialogue. Tasks, therefore, are central in direct tests. They are used as testing frameworks that set the context. The FSI-type oral interview is 'the earliest and most influential of all approaches' (Fulcher, 2003: 173) to testing speaking. While it was initially used for military purposes, it was adopted by many universities and schools in the 1970s (Fulcher, 2003). It became the basis for curriculum designs of modern foreign language teaching in the U.S (Fulcher, 2003).

However, despite its wide use and its long history, the OPI has been heavily criticised partly because its validity lacked empirical evidence.

Another type of direct testing, which was introduced by the University of Cambridge Local Examinations Syndicate (UCLES) in the 1990s, was *testing in pairs or groups.* This involved two or more test-takers interacting with an interlocutor/examiner. The First Certificate in English examination, which implemented the paired format, contained an interview, an individual long-turn for each test taker, a two-way collaborative task and a three-way discussion between the two test-takers and the interlocutor (Fulcher, 2003). However, through a large body of empirical as well as theoretical research on the paired or group testing approach, a debate among researchers emerged.

This body of research highlighted both the strengths and the weaknesses of this approach (Taylor & Wigglesworth, 2009). Several advantages were found in the use of the paired format for testing speaking. In a number of unpublished studies, the UCLES found that in this format, the length of the candidate's turns was increased, while the assessor spoke for substantially less time (Fulcher, 2003). Among its advantages was that it provided the opportunity to test a richer and more complex output than in the individual format (Fulcher, 2003). It was also seen as a better reflection of a typical classroom (Bonk & Ockey, 2003; Hilsdon, 1991), resulting in less anxiety for the test takers. In a recent empirical study at a Canadian university, Brooks (2009) compared the performance of 16 participants in the individual format (interacting with an examiner) and in a paired format (interacted with another test taker), and she found that the participants performed better in the paired format. Analysing the data within Vygotsky's SCT theory, she further reported that the interaction of test takers in the paired format resulted in more complex output, more negotiation of meaning and more consideration of the interlocutor.

However, the paired format had not always been supported, as it raised several issues. Some researchers found it problematic because it made it difficult to assign an individual score to each test taker (Foot, 1999; Fulcher, 2003; Norton, 2005). Another crucial issue was the effect of the interlocutor. Although interaction with an interlocutor features in the individual format as well, the variability of the interlocutor effect in the paired format is greater (Brooks, 2009). Using data from Japanese students taking the Cambridge Speaking tests, Norton (2005) argued that the interlocutor's linguistic ability, gender and the extent to which they are familiar with each other could affect the amount of output and involvement in the test. In terms of the potential effect of the personality of the interlocutor, no evidence was found suggesting that this effect would be greater in paired formats than in individual formats of testing (Van Moere, 2006).

The second type of speaking tests, namely *indirect tests*, is conducted with the help of computers or tapes and they involve no human interaction. The test taker is given a prompt to speak to a microphone. Examples of this test are the simulated oral proficiency interview (SOPI), the PhonePass test (Ordinate) and the Test of Spoken English (Educational Testing Service) (Fulcher, 2003: 190). Some researchers (e.g. Stansfield & Kenyon, 1992) favour indirect tests because the absence of the human element minimises the possibilities of confounding variables to occur, and consequently ensures reliability. Other researchers (e.g. McNamara, 1997) however, condemn indirect tests for the same reason. For them, 'the co-construction of meaning in discourse is seen as essential to communication' (Fulcher, 2003: 190), and this can only be achieved with human interaction.

As there is no consensus on a speaking test that is the most valid and reliable, a test that was suitable for this kind of participants and for the purposes of the study was selected. Due to all the variables that arise in paired format testing, a direct test with one test taker was preferred

in the present study. A structure similar to the oral test of the *International English Language Testing System (IELTS)* (e.g. Issit, 2008) was adopted, in order to make the test fairly standardised. The test consisted of two parts: a *Question/Answer* part and a *Presentation* part. The Question/Answer part tested *interactional fluency,* by engaging students in a short conversation with me. A list of questions was prepared for this part, in order to make sure that all students received the same questions. An effort was made to order these questions in a more open-ended sequence as the conversation unfolded, so that the level of the students' competence would be revealed. The Presentation part tested *unidirectional fluency,* as it required students to engage in a monologue. Students were given a minute to prepare to talk about a person they were familiar with. Some short questions were provided as prompts to help the participants build up their presentation. During their preparation minute, they were allowed to take notes to help them remember what they wanted to say.

When designing the tests, the main consideration was for these not to resemble the types of tasks used during the intervention, in order not to favour the EO group who had the experience of working under L2-only conditions. Moreover, by having a different test to the intervention, the effect of repetition was avoided. In other words, the experience the students gained on certain types of tasks during the intervention would not favour them during the tests. In addition, having different tests minimises the possibility of the students losing interest in their activities.

Although the three tests had the same structure, their topics were different, again in order to eliminate the repetition effect. What was important in selecting the topics of the oral tests was *topic familiarity*. As there is evidence suggesting that prior knowledge plays an important role in oral and written performances (e.g. Pulido, 2004; Leeser, 2007), students were asked about personal matters. As shown in Appendix 11, for the pre-test, students were asked about their

favourite hobby and they presented their favourite singer or group. The first post-test (Appendix 12) asked questions about their favourite film or their favourite TV series, while they prepared a short presentation about their best friend. Finally, the delayed post-test (Appendix 13) asked about their best holidays, while presenting a person they admired.

All three tests were conducted by me and they took place in a quiet room outside the classroom, in which only the student and I were present. Each test took approximately ten minutes to be completed for each student, during which the students were audio-recorded.

3.5.4. Fidelity to conditions

One of the main challenges of the design was *fidelity to conditions.* It was crucial for the participants of all three groups to work under the language conditions imposed. Regarding the CS group, fidelity to conditions was not difficult to be preserved, as students generally tend to switch to the L1 when they speak amongst themselves. At the beginning of every lesson, they were reminded that they should not hesitate to switch to Greek when they need to, while being recorded. As for the COM group, fidelity to conditions would be preserved, as long as the participants of this group completed the tasks once a week. As I was there on every lesson observing and recording them, I can confirm that fidelity to conditions was preserved in this group as well.

The difficulty with preserving the conditions was with the EO group, who were not used to speaking in English amongst themselves at all times. In order to keep their use of Greek to a minimum, I kept stressing that they should make an effort to stick to English even when they speak about how to solve a task. I clarified what I meant by giving them examples of the instances they could use English for. Furthermore, during the completion of the task, I was

walking around listening to their interactions, reminding them of the English-only policy where necessary. Realistically, as long as this group kept the L1 to a minimum in relation to the CS group, then the conditions would be preserved.

For this reason, three pairs from each group from four weeks of the intervention (Week 1, 3, 5 and 7) were selected based on their representativeness. This means that the pairs who were preferred were the ones who carefully listened to my instructions before starting the task (recorders started as soon as I started giving instructions, so I am able to say which pairs listened to me before attempting to complete the task); pairs who remained with the same interlocutor for both sessions of that week, as well as for most of the weeks of the intervention; and pairs who completed the whole task. Based on the transcripts, the amount of L1 use was found by calculating the percentage of L1 use for each group for each week. Figure 3.5.1 below presents the results.

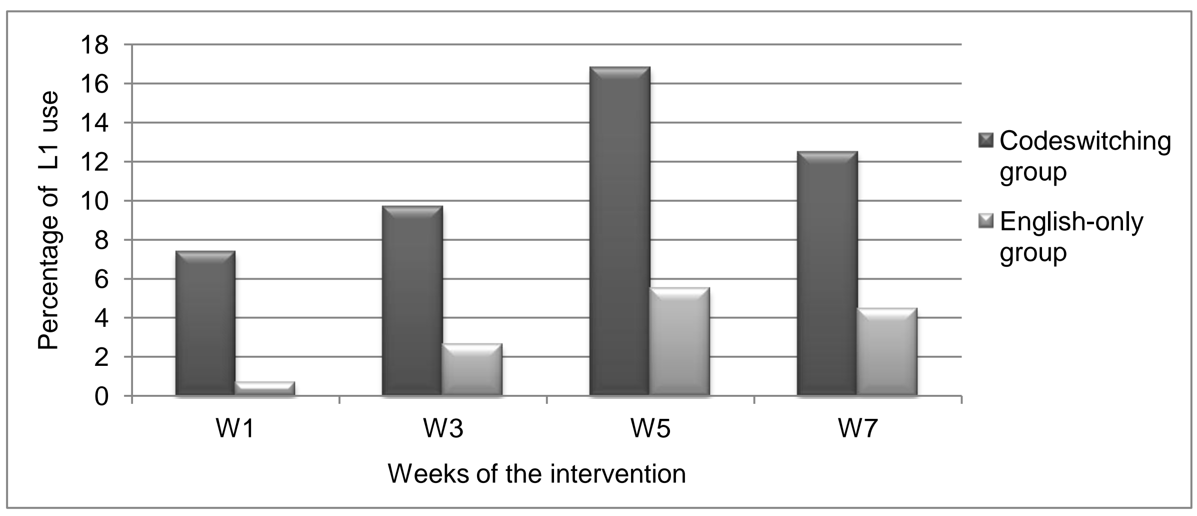


Figure 3.5.1 Percentage of L1 used by the two experimental groups on four weeks

As shown in the figure above, overall the EO group switched a lot less to the L1 than the CS group on all four weeks. The highest percentage of L1 used by the EO group was less than 6% (Week 5). As for the CS group, the lowest percentage of L1 use was approximately 8%. It is clear therefore that in relation to the CS group, the EO group did keep their L1 use to a

minimum and it could be argued therefore that fidelity to conditions was preserved with the EO group as well.

Another factor that was considered for fidelity to conditions was the *teacher variable.* Two teachers participated in the study. Teacher A was teaching English at the two schools acting as the experimental groups, while Teacher B was teaching English at the sixth grade classes of the school acting as the comparison group. In order to ensure that their teaching behaviour would not act as a confounding variable, a total of twelve lessons were observed and audio-recorded during three periods of time: the pilot study, a few weeks before the main study and during the main study. The reason these recordings took place at several time points was twofold: first, it would ensure that the teachers were similar before the intervention and kept being similar during the intervention; and second, it would help avoid Labov's (1972) *observer's paradox,* according to which the person under observation is not behaving naturally because s/he is aware of being observed. It is worth mentioning here the discomfort that these teachers felt every time I asked to record them. Despite the fact that this was pre-arranged, they would sometimes ask me to postpone the recording, as they were not feeling comfortable with being recorded that day. An effort was made to observe lessons on the same unit of the book so that the teachers would be more comparable.

Regarding the first purpose of observations and recordings, that is to ensure that the teacher variable did not affect the situation, the teachers were monitored in terms of the distribution of their L1/L2 speech during classroom time and the functions that their L1 use served. Having listened to all the recordings, I ensured that overall the teachers were using the same amount of L1 use between them and in all their lessons. More specifically, the teachers spoke mainly in the L2 and switched to the L1 to serve some kind of purpose. In order to confirm this observation, I transcribed one lesson from each teacher that took place during the

intervention and calculated the percentage of L1 use during those two lessons. The result which appears in the table below confirmed that the amount of L1 use of the two teachers was very similar. Both teachers used less than 8% of Greek during their lesson.

Table 3.5.2 L1/L2 distribution of two teachers

Teacher A	Teacher B
Total words used: 1565 L1 words used: 106	Total words used: 1507 L1 words used: 107
Percentage of L1 use: **6.77%**	Percentage of L1 use: **7.10%**

Looking at the functions of their L1 use, it seems that overall the two teachers did not differ very much in that respect either. In the excerpts of transcripts appearing in appendices 14 and 15, the utterances in bold highlight the teachers' switches to Greek and they are followed by an English translation. All students' names that appear in the transcripts are pseudonyms, which were used in order to preserve the anonymity of the students.

One of the most common L1 functions used by both teachers was vocabulary translations (lines 36, 48, 80 for teacher A; lines 17, 87, 91 for teacher B). The teachers used the L1 equivalent of words, or asked the students to give it to them, when they encountered a less frequently occurring L2 word. In addition, they seemed to be using the L1 for encouragement. They tended to switch to Greek in order to motivate students to speak (lines 110, 112 for teacher A; lines 3, 27, 35 for teacher B). Moreover, although instructions were generally given in the L2, at some points the teachers switched to the L1 for instructions, in order to ensure that all students understood what they had to do. Such switches appear in teacher A's transcript on line 100, and in teacher B's transcripts on line 39. Finally, teacher A used the L1 in order to draw the attention of her students to certain irregularities of the forms

at several points during her lesson (lines 36, 64, 80, 114). Teacher B expressed her surprise in Greek when the students were not familiar with a famous athlete on line 5.

The fact that the two teachers have almost the same L1/L2 distribution in their lessons and their switches serve almost the same functions could be considered as proof for eliminating the possibility of the teacher variable affecting the situation of the intervention.

3.6. Ethical considerations

Having received ethical clearance from the University of Oxford Central University Research Ethics Committee, as well as the Ministry of Education in Cyprus, three schools have been contacted. After an initial oral conversation with the headteachers of the schools, each headteacher received a different letter (Appendix 16) explaining the study requirements in detail. This was accompanied by a consent form (Appendix 17), which the headteachers completed if they agreed for their school to participate in the study.

Then, the English teachers of the sixth grade were approached. Three different letters (Appendix 18) were prepared for the teachers, each describing a different condition. With the teachers' written consent (Appendix 19), I then visited all sixth grade classes of these schools to introduce myself and explain what they would be doing in case they, and their parents, agreed to participate in my study. All students were given an information pamphlet for their parents or guardians (Appendices 20, 21, 22), accompanied by a written consent form (Appendix 23) for the parents to complete, if they agreed for their children to participate.

In all the aforementioned documents, I briefly introduced myself, while the purpose and the procedures of the study were explained. The documents also stressed that the anonymity of all the participants would be preserved and that confidentiality would be kept at all levels.

Finally, participants were reminded that they had the right to withdraw from the study, if they wished to, at any point without any consequences.

Finally, the students with parental consent were given a slip (Appendix 24), which asked them to confirm that, apart from their parents, they also wished to participate in the study. All ethical procedures described for the main study were followed for the pilot study as well. At the end of the main study, all students participating received a small gift (University of Oxford pen) as a gesture of gratitude.

3.7. Pilot study

A pilot study took place in October 2010. It lasted for one month and it served three main purposes: first, to trial the two experimental conditions; second, to try out different types of tasks; and third, to pilot the type of oral production test that would be used in the main study. One class of thirteen sixth grade students of a primary school in Nicosia participated in the pilot study. This school was not one of the three schools which participated in the main study.

3.7.1. Piloting the experimental conditions

i. English-only condition

During the first week of the pilot study, the English-only condition was trialled. The lesson was taught as normal and, during the last ten minutes of the lesson, the students were asked to complete a task with the person sitting next to them. Three types of tasks were prepared for this lesson, so each task was completed by two pairs (one was a group of three). The tasks

were an information gap task, a role play and a 'spot the differences' task, which were designed based on the structure taught that day, namely the superlative form.

Despite the English-only conditions, the L1 was mainly used for metalanguage. In terms of *metacognitive talk*, that is talk about the task, the students used Greek to set the goals of the task, assign roles, decide what to do next, determine turn taking and comment on their performance. Greek was also used for *metatalk*, that is talk about the language, which included identifying sources of linguistic difficulty, such as vocabulary difficulties (see examples in Appendix 25).

However, apart from metalanguage, a few students expressed some task language words in Greek. As Appendix 25 shows, these were nouns and they were the Greek equivalents for the words 'tail', 'Russia', 'one hundred and twenty' and 'thirty-three'. Most of the students in the same classroom did not find it difficult to use the English equivalents of these words. However, a small number of students were unable to express these words in English. This suggested that student proficiency played an important role in the study. The final research question therefore pursued this issue.

ii. Codeswitching condition

During the second week of the pilot study, the same procedures were followed under the conditions of the CS group. The tasks were an information gap task, an opinion gap task and a role play task, which focused on making suggestions. The students were allowed to use Greek whenever they felt that they could not express themselves in English. Similarly to the previous week, students used Greek for both metalanguage and task language (see Appendix 26). The fact that the L1 was used for task language by some students supports the argument for proficiency discrepancies. During the pilot, a handout with feedback was given out to the

students. The recordings of the repeat lesson showed that the *recycling process* was successful and that all students used the feedback.

A final comment on the experimental conditions concerns the recording variable. It seems that the students felt that what was being recorded should be flawless. Even in the CS condition, students were making an effort to complete the tasks in English only. Also, students resorted to whispering for metalanguage because this was in Greek and they thought it was inappropriate to for this to be recorded. Third, at several instances, the students asked me to re-start the recording because, as they said, they performed badly. All these indications suggested that audio-recording could act as a confounding variable in the study. Nevertheless, it was expected that in the main study this variable would be minimised because the students would get used to being recorded and would feel more comfortable as the time went by.

3.7.2 Piloting tasks

One of the primary aims of the pilot study was to try out different types of tasks in order to determine the most suitable for the purposes of this study and the kind of sample. Four types of tasks were trialled throughout the pilot study: information-gap tasks, role plays, a ‘spot the differences’ task and an opinion gap task. The conclusions reached in this section were based on the recordings and the data collected from short interviews with seven of the students. The students were asked about the difficulty of the tasks, the necessity of L1 use during some tasks, and whether the tasks were fun to complete.

In order to determine whether the tasks were appropriate for the proficiency level of the students, two indicators were considered. First, the students’ understanding of the task instructions was considered. For all types of tasks, the students asked me to explain what they had to do. Some students simply wanted to confirm that they understood, while others needed

further explanations in Greek. Having informally talked to the teachers about this, it seems that this is normal practice. Greek explanations almost always follow English instructions of tasks or other exercises, mainly to ensure that everyone understood. Therefore, the fact that almost all students asked for Greek instructions should not be considered as an indicator of inappropriate level of task difficulty. First language explanations were added thus, to the procedures that I would follow in the main study.

Moreover, the quantity of L1 used by the students while completing the tasks was a good indicator of task difficulty. Quantity of L1 use could be directly linked to three other variables: the *experimental conditions*, *student proficiency* and *task types*. First, students were more comfortable using the L1 under the CS condition, than under the EO condition. Second, more proficient students were able to complete the tasks without switching codes, despite the condition, while less proficient students struggled to express simple words in the TL. Third, task type influenced L1 quantity. Tasks with mainly visual input and almost no L2 cues other than the example (e.g. Spot the differences) increased the students' use of Greek. However, tasks with lots of L2 cues (e.g. opinion-gap task) were easier to complete in the L2 only. Students simply put together parts of language that were provided somewhere in the task to structure their sentences, rather than being creative. Tasks with too many L2 cues, therefore, were avoided in the main study.

In general, it was concluded that the tasks were appropriate for this type of sample and for the purposes of this study. What was taken into consideration when designing the tasks of the main study was reducing the L2 cues, while increasing the visual input. As the interview data suggested, the students enjoyed completing these tasks and they did not find them extremely difficult, so the same types of tasks were used in the main study.

3.7.3 Piloting the oral production test

During the pilot study, the oral production test was trialled. The structure of the test was as described in section 3.5.3, but with an additional section. This was a role play task, set at a travel agency where the client, played by myself, asked the employee, namely the student, questions about Cyprus. This part was subsequently deleted because it did not offer students the opportunity to create their own language, and because it looked too much like the role-play tasks given in the intervention.

Seven students from the classroom participating in the pilot study were tested separately and recorded. The test did not take more than ten minutes to complete. All students seemed enthusiastic about the topics, as they were all keen on talking about their favourite hobbies, singers or groups. In general, the test seemed to be appropriate for measuring fluency because it allowed learners to produce some kind of a monologue, while both parts of the test were very engaging.

3.8. Matching participants

As discussed earlier in this chapter, one way of strengthening a quasi-experimental design is by *matching participants,* which makes the groups more comparable. This process, which is performed on a case-by-case basis, was based on four possible confounding variables, discussed in this section. The results are presented in the Quantitative Results chapter.

3.8.1. Second Language Proficiency

The first variable is *student L2 proficiency*. Measuring L2 proficiency serves two purposes: first, to ensure that all groups have a similar average level of proficiency; and second, to investigate the third research question, which looks at the benefits of TR+ on different

proficiency groups. The C-test was chosen as a measure of proficiency because it is widely used in various contexts, such as schools, universities or workplaces; and for different purposes, for example as a placement test or as an instrument in cognitive and applied linguistics research (Eckes & Grotjahn, 2006). With such wide use therefore, the C-test has established itself in the field of language testing and assessment (Eckes & Grotjahn, 2006) as a measure of general language proficiency.

C-tests, which are the newer version of cloze tests, are the operationalization of the *reduced redundancy principle* (Spolsky, 1973). The rationale is that languages are redundant by nature, so speakers of the language should be able to supply the missing linguistic items under such conditions (Babaii & Ansary, 2001). The test, therefore, measures the abilities of the examinee to comprehend a linguistic message 'with some noise or interference' (Lee-Ellis, 2009: 247) in a metaphorical sense. C-tests usually consist of four to six short authentic texts, preferably placed in order of increasing difficulty (Klein-Braley, 1997). In order to apply the reduced redundancy principle, the *rule of two* is applied (Klein-Braley & Raatz, 1984). The second half of every second word is deleted, starting with the second sentence of each text (Klein-Braley, 1997). The test takers are asked to complete the missing parts of the words.

The C-test has three very appealing advantages. First, it is highly reliable (Eckes & Grotjahn, 2006). Second, it is a very economical measurement instrument, as it is easy to administer. And third, scoring C-tests is fast, easy and, most importantly, very objective (Klein-Braley & Raatz, 1984). Despite these advantages, C-tests have been criticised for their lack of face validity (Bradshaw, 1990; Jafarpur, 1995) and their poor item discrimination (e.g. Cleary, 1988). Nevertheless, there is evidence suggesting that C-tests measure the same variable as most language assessment tests (Lee-Ellis, 2009). C-test scores are highly correlated with scores of other types of institutionalised proficiency scores, such as 'the TOEFL (r = .55 to

.91), the TOEIC (*r* = .62), the Michigan Test (*r* = .54 to .61) and the Oxford Placement Test (*r* = .83)' (Lee-Ellis, 2009: 248).

Having in mind the purpose of the present study, the high practicability and validity of the C-test provide good reasons to select this test as a proficiency measure. A C-test was designed based on four passages from four different units of the current English book (Appendix 27). As most of the units introduce language through dialogues, these dialogues were turned into short passages through some minor modifications. Having applied the rule of deletion, a total of 91 missing gaps were created. The C-test was piloted with 12 sixth grade students from the primary school that participated in the pilot study. Half of the students scored very high (above 80 out of a 100), while others did not perform so well. It was concluded therefore that the test was suitable as it reflected the variability of proficiency level in that class.

In the main study, the C-test was administered a week before the pre-test. The students were given fifteen minutes to complete it individually. The tests were scored by me. Two points were given to every correct answer with no spelling mistakes, one point to every misspelled correct answer and no points were given to wrong answers or blank gaps. Due to its very objective scoring system, no inter-rater reliability was sought. The total score for every student was calculated against 100, instead of 182 (91 gaps x 2 points). A split from the median divided the students into two groups: the high proficiency group *(HP group)* and the low proficiency group *(LP group).* The results for each group are presented in Chapter 6.

3.8.2. Productive Vocabulary

Productive vocabulary, or in other words, the range of L2 words that are active and are being used (Nation, 2001), was also measured prior to the intervention. As the students were asked to produce language both during the intervention and the oral production tests, it seemed reasonable to measure this variable prior to the study. Productive vocabulary is very difficult to measure, as it 'tends to be so context-specific that it is difficult to calculate from any small sample the true size or range of the learner's productive vocabulary' (Meara & Fitzpatrick, 2000: 20). A number of different types of tests were used in the past, which can be divided into three major categories: *Controlled Productive Vocabulary tests, Free Production Vocabulary tests* and *Word Association tests.*

In *controlled productive vocabulary tests* target words are predetermined. Testees are presented with 'a sentence context, a definition, and/or the beginning of the target word' (Meara & Fitzpatrick, 2000: 20) and they are required to complete the missing word. An example of a controlled test is a *translation test* (used in Webb, 2008), which presents a list of L1 words and requires the L2 equivalents. Another example is *elicitation tests,* such as C-tests or gap-filled tests, which provide sentences with some incomplete words (often the first one or two letters given). Advantages of these tests are their quick and easy construction and marking system (Milton, 2009), as well as their reliability. Controlled tests had been criticised for being mainly effective with lower level learners who have a limited vocabulary size (Meara & Fitzpatrick, 2000). Nevertheless, this disadvantage is eliminated in the present study, as the participants are lower level learners.

Free productive vocabulary tests set up a situation where the testee produces either spoken or written discourse. In analysing the output, the vocabulary is categorised into 'frequent, less frequent and infrequent words' (Meara & Fitzpatrick, 2000: 21). The more infrequent words

the testee produces, the larger his/her productive vocabulary is. These tests have been criticised because they require a large amount of production, in order for infrequent words to occur in a grammatical and well-structured environment (Milton, 2009; Meara & Fitzpatrick, 2000). As learners often tend not to produce large quantities of language, free production tests do not always yield solid results.

Lexical diversity, or the variety of vocabulary used by testees (Malvern & Richards, 2002), is often used as a measure of this type of tests (Milton, 2009). This is often represented by the *Type Token Ratio (TTR)* (Johnson, 1944), which is calculated by dividing the number of *types,* that is the number of different words used, by the number of *tokens,* that is the total number of words used (Richards & Malvern, 1997). The higher the score, the larger the lexical variety, while the lower the score, the more repetition occurs. Despite its wide use, the TTR has been criticised for being unreliable (e.g. Broeder *et al.*, 1993; Vermeer, 2000) because first, it does not consider whether words are high or low frequency; and second, it is mathematically flawed because the assumption that it is independent of sample size is false (Richards & Malvern, 1997). As sample size gets larger, the TTR score significantly decreases because the speaker's active vocabulary is exhausted.

Finally, *word association tests* give testees stimulus words and ask them to produce a word in response. An example of word association tests is Meara and Fitzpatrick's (2000) *Lex30*, which presents testees with 30 stimulus words and asks them to produce four association words for each one. According to Milton (2009) however, the problem with word association tests is that once testees realise that the aim is to produce infrequent words, they might produce any infrequent words which are not necessarily associated with the stimuli words. It becomes easy, therefore, to get a high score and difficult to decide which words are acceptable associations.

From the discussion above, it is clear that there is no generally ‘accepted, standardised method for measuring L2 productive knowledge’ (Milton, 2009: 146). For these participants therefore, a translation test was selected as the most suitable, due to its reliability, its comprehensibility and its ease of completion and scoring. A list of thirty L2 words of different word classes were selected from the students’ syllabus (Appendix 28). These words were selected from the units that had already been covered before the day of the test. The aim was for this list to include different types of words (i.e. nouns, adjectives, verbs). Moreover, I tried to include words that were more likely to occur on a regular basis and therefore more students were likely to know (e.g. “teacher”, “yesterday”), as well as words that were only encountered once during the recently covered units and students would be less likely to remember them (e.g. “rock”, “capital”). This range of words would be able to distinguish between students with higher and lower productive vocabulary. The students were presented with the L1 equivalents of these words and they were given fifteen minutes to produce the L2 target words. The test was piloted with the twelve students participating in the pilot study. As the results reflected a range of productive vocabulary ability, the test was judged suitable for the purposes of the study.

The scoring system followed for the C-test was followed for this test as well. As this is an objective system, no inter-rater reliability was sought. Based on a median split, two categories were formed: high productive vocabulary group *(HPV group)* and low productive vocabulary group *(LPV group)*.

3.8.3. Socio-economic status (SES)

Socioeconomic status, which is a variable that determines the social background of an individual, has been directly linked to the overall L2 achievement of learners in education research (Ellis, 2008). Evidence suggests that learners from lower socioeconomic groups are not as successful educationally as those coming from upper socioeconomic groups (e.g. Burstall, 1975; Skehan, 1990). In terms of the present study, SES is potentially a significant factor, as students in this context attend private English lessons after school. Students of a higher SES background therefore are more likely to have received longer or better education in private English institutions, than students of a lower SES background. SES therefore is of particular importance in the present study.

A short questionnaire (Appendix 29) was prepared for the parents of the students to complete, in order to collect data on the family's SES. The questionnaire was included in the envelope, which was sent to the parents of all students, along with a parents' pamphlet and consent form. Having mentioned the importance of SES in the study and having stressed the anonymity of the data once again, the parents were asked to complete this questionnaire, if they wished for their child to participate. The questionnaire covered three main areas, which contribute to SES: parental educational level, parental occupational status and family income. In all three sections, a list of possible answers was given and the parents had to tick the boxes which applied to their household.

A two-rank SES classification was formed: *HighSES* and *LowSES*. In order to allocate the participants in one of these two groups based on the responses of the questionnaire, numbers were assigned to each response and a cumulative procedure was followed. For the first part of the questionnaire, which represented the parental educational level, no points were given for the "Did not attend school" option, one point for the "Primary school certificate", two points

for the “Secondary school certificate” option and so on. The maximum points that a student could be given from this section was ten (five for each parent with a postgraduate degree).

For the second section, the points given were either one or zero. When a parent had some kind of occupation, either in the public or private sector, or as a self-employer, one point was given. When a parent did not have an occupation, either because s/he was a pensioner or because s/he was unemployed, no points were given. The procedure followed for section A was also followed for section C. For an income of less than €2000 per month, zero points were given. For an income between €2000 and €4000, one point was given and so on. The highest score that a participant could get from this section was four, which represented an income of €8000 or above.

The scores of all sections were then added and a final score was given to each student. Participants who scored above the median were allocated to the *HighSES* group, and students who scored lower than the median score were allocated to the *LowSES* group.

3.8.4. Prior contact with the English language

The vast majority of the students in Cyprus attend private English lessons in the afternoons from a very young age. It is highly unlikely to find a student at the age of 12 who received no private lessons of English. As previously discussed, this is linked to the SES variable, as students of a higher SES are more likely to have had more extensive contact with the English language. An attempt was made therefore to assess this prior contact of the participants with the English language. A short questionnaire (Appendix 30) was prepared and given to the students to complete. In order to determine whether the respondent had some kind of prior contact with English, questions about living abroad, having foreign parents or having any

other type of contact with English (e.g. English-speaking kindergarten) were included. The respondents were also asked about whether they had been taking English private lessons and, if they did, for how many years.

This *contact* variable was treated as an ordinal variable. The participants whose only contact with the English language was at school were allocated in the lowest category. The second category, namely *Private lessons*, included the students who had up to five years of private English lessons, as this was found to be the most common. The following category, *Private lessons + Extra years*, included the students who had more than five years of private English lessons or attended an English-speaking kindergarten. Finally, students who had an English-speaking parent or parents, plus the private lessons, were allocated in the *Private lessons + Extra years + Parents category*.

3.9. Other possible confounding variables

3.9.1. Researcher's protocol

Having controlled for as many variables as possible, the only variable that differed among the three groups were the conditions under which the tasks were completed. As I was the sole person responsible for ensuring that the conditions were carefully implemented, a protocol (Appendix 31) was prepared that I followed in every lesson of every group. This protocol was a step-by-step guide which provided a detailed description of what I had to do in every lesson of each condition. It was a way to ensure that no group would be favoured in any way.

3.9.2. Recording Factor

The audio-recording method was central to the data collection process, as it was used for the students' interactions, the oral production tests, the recordings of lessons and, as it will be discussed later, for the recordings of teacher interviews. It was vastly present because it is the only way to ensure that all available data are recorded without any information, which may initially seem irrelevant, being omitted.

The problem with this method is that it may make the participants become more cautious about what they say. This is linked to the *observer's paradox* mentioned earlier, which refers to cases of people behaving abnormally under observations. Indeed in the present study, metalanguge was often whispered, as the students felt that this off-record language should not be recorded. In order to minimise this effect, students of all groups were reminded that I would be the only person listening to their recordings, so they should feel comfortable and refrain from whispering. My impression is that because these students were being recorded for a relatively long period of time, they started getting used to the recorder and whisperings were gradually minimised.

3.10. Teacher Interviews

After the end of the intervention, interviews with English teachers of different urban primary schools were conducted. The purpose of these interviews was to gain a more profound understanding of the role that task-based learning currently has in these classrooms and investigate its potential within this context. The interview data, which are complementary to the main study, are analysed qualitatively in response to the first research question.

3.10.1. Sampling procedure – Ethical considerations

The *criterion sampling* procedure was followed for the selection of teacher participants. According to this method, teachers who meet certain predetermined criteria can be recruited (Dörnyei, 2007). Two criteria were set for the selection of the teachers: first, currently teaching English to sixth grade students; and second, teaching in urban schools in Nicosia.

From a list of all primary schools in urban Nicosia, I randomly called 19 schools and I asked to speak to the teacher, who teaches English at sixth grade students. Having explained to the teachers what the purpose of the study was and what they would be asked to do if they agreed to participate, we then arranged a meeting at the school. At the meeting, the teachers were given a letter (Appendix 32) with a more detailed description of the purpose of the study, my contact details, as well as a consent form (Appendix 23). With the participants' consent, an appointment for the interview was arranged at a convenient time and place for the teachers. A total of 21 teachers were interviewed, including the two teachers who participated in the main study. One of the Inspectors of English in Primary Education at the Ministry of Education and Culture of Cyprus was also interviewed. The same interview schedule was adapted accordingly for the Inspector's interview. This is the interview from which the information in the Introduction chapter came from.

3.10.2. The Interviews

The interviews were *semi-structured.* Although an interview schedule was prepared in advance with certain broad questions, this had a guiding role, as the interviewees had the opportunity to elaborate on any issues they found to be relevant to our discussion (Dörnyei, 2007). All interviewees were asked the same questions, but not necessarily in the same wording or the same order. This made the answers more comparable across the respondents.

All interviews were recorded with a small recording device. Although the recording aspect did pose a concern to some of the interviewees, it was decided that it was important to record the interviews, as note-taking is not an adequate method of recording rich data. Concerned interviewees were re-assured that the recording would be anonymous and that the data would not become available to anyone but me.

3.10.3. Interview Schedule

An interview schedule, which included questions emerging from the review of the literature and from the present inquiry, was prepared. This interview schedule was first used for the interviews of the two teachers whose classes participated in the intervention study. These initial interviews acted mainly as pilot interviews, that would not only pilot the interview schedule, but also reveal other issues that are worth investigating further.

Based on the themes that emerged from the first two interviews, a finalised interview schedule was prepared (Appendix 33). The questions were divided into four major categories, which broadly concerned task implementation, task repetition, metalanguage and the role of pairs during task completion. A few questions were prepared for each category, but the interviewee was free to switch to a subject that they thought was relevant, but was not directly asked.

3.11. Data Analysis

This section explains the data analysis procedure followed for each research question.

3.11.1. Analysis for Research Question 1

The first research question examined the perceptions of teachers in Cyprus on matters relevant to task-based learning, as well as to the intervention. The data for this question came from the recordings of teacher interviews. These data were analysed qualitatively in relation to the four topics set for the interview discussions: task implementation in Cyprus, task repetition, metalanguage and the role of proficiency pairs during task completion. The qualitative analysis of the interview data are presented in Chapter 4.

3.11.2. Analysis for Research Question 2

The second research question examined whether the content of the students' interactions changed in any way from session 1 (S1) to session 2 (S2). The data came from the recordings of the students' interactions in the two experimental groups during the eight weeks of the intervention; the comparison group did not repeat any tasks and was therefore not considered in this question.

A qualitative approach was adopted for the analysis of these data. Having transcribed a good proportion of the interactions (60%-70%), the S1 interactions were compared to those of S2 for each week. The main focus of the analysis was to identify whether task language and L2 metalanguage evolve in any way from S1 to S2. Reference was also made to the role of L1 metalanguage, scaffolding and the effect of accuracy feedback where appropriate. In addition,

the conditions under which the interactions took place were taken into consideration while analysing the results. The results of the qualitative analysis of the interaction data are presented in Chapter 5.

3.11.3. Analysis for Research Question 3

For the last three research questions, which concern the outcomes of the intervention, the data came from the three oral production tests (pre- & post-tests). These were transcribed and analysed quantitatively. The transcription was completed with the help of *NVivo;* a data-analysis software which reduced the speed of the recording up to 50%. All data from the tests were entered into SPSS, as this was the software used for all statistical tests. The results of the quantitative analysis of all three research questions are presented in Chapter 6.

In terms of the third research question, the aim was to investigate whether TR+ leads to more WTC than no task repetition. It further investigated the first part of the hypothesis, which predicted that task repetition with codeswitching *plus* would lead to more WTC than task repetition with English-only *plus*. As discussed in the literature review chapter, self-report measures of WTC, which had been very popular in past research, were not chosen for the present study, as they reflect perceptions on WTC, rather than actual WTC levels. In terms of my study in particular, asking such young learners to complete questionnaires on how willing they are to speak at certain moments was judged to be an inadequate measure.

As discussed in the previous chapter, based on McCroskey's (1992) WTC scale which was relevant to length of time, WTC can also be conceptualised as expansion of speech, rather than just initiating communication. Considering the type of oral production tests that were used as pre- and post-tests, the conceptualisation of WTC as expansion of speech made more sense for the present study. For the oral production tests, participants were asked pre-

determined questions in the Question/Answer part, and presented a monologue in the Presentation part. Not much space was given therefore for them to show their WTC through turn taking; sustaining communication was a more appropriate measure.

In order to adequately measure the expansion of oral production, *speaking time* was chosen as the first measure of WTC. This was measured in seconds and it represented the time spent by the student speaking or pausing. Both parts of the oral production test were considered for this variable. Nevertheless, for the Question/Answer part, the time allocated to the researcher posing the questions was deleted. The calculation of the speaking time variable was completed with the help of NVivo. By reducing the speed of the recording, it became easy to calculate time. In addition, the *L2 syllables* variable was chosen, as expanding oral production also means producing more output. This measure represented the total number of L2 syllables produced by the student during the whole test and they were counted manually by me.

For reliability purposes, roughly 10% of the sample (ten students) was selected for re-analysis based on the *systematic sampling method.* According to this method, every 7th name from the list of all participants was selected. For each student, all three oral production tests were re-analysed; that is a total of thirty tests. The re-analysis for these variables was conducted by myself. Intra-rater reliability was judged to be suitable for these variables, as speaking time and the total number of L2 syllables, are quite objectives measures. In other words, the personal judgement of the rater was irrelevant for these measures. A Pearson's correlation was conducted for each variable between the first and the second set of data. The results, shown in Table 3.11.1 below, indicate a very high correlation between the two sets of data.

Table 3.11.1 Correlations for intra-rater reliability: WTC measures

WTC measures	Correlations *(r)*
Speaking time	.99**
L2 syllables	1**

**p < .01

3.11.4. Analysis for Research Question 4

The fourth research question aimed to examine whether TR+ has an effect on fluency development. It also inquired about task repetition with codeswitching *plus* being more effective than task repetition with English-only *plus*, in terms of oral fluency development. The question assumed that this development would come with no detrimental effects on accuracy. In order to fully respond to the question, three variable constructs were formed: *fluency variables, content variables* and *accuracy variables.*

i. Fluency variables

As discussed in the literature review, fluency has two components: a temporal component and a dysfluency component. Three measures were chosen for each component based on measures used in the past. In terms of the temporal component, two of the best predictors of fluency, namely *speech rate* and *mean length of runs* (Kormos, 2006), were chosen. *Speech rate*, measured with syllables per second (Towell, Hawkins & Bazergui, 1996), was calculated by dividing the total number of syllables by the total number of speaking time measured in seconds. *Mean length of runs (MLR)* is the average length of runs that occurs

between short pauses. This was calculated by dividing the total number of syllables by the number of runs, which are the utterances between short silent pauses. The cut-off point for pauses is discussed next in this section. The total number of syllables used for the calculation of both measures included both L1 and L2 syllables. This decision was taken in order not to intervene with the data. For this reason, the L1 syllables measure was also considered when interpreting the results.

Pause rate, that is the percentage of time spent in both filled and unfilled pauses against the total speaking time, has also been widely used (e.g. Lennon, 1990; Temple, 2000). This measure was preferred to silent ratio, which considers time spent in unfilled pauses only, because time spent in both filled and unfilled pauses is considered to be time spent thinking. Filled pauses were pauses with sounds (e.g. "eh", "uh"). While there is no consensus among researchers about the cut-off point that should be used for pauses in speech analysis (Towell *et al.,* 1996), most researchers used a cut-off point of .25 seconds (Kormos, 2006). However, since most studies on fluency recruited adult participants, it was decided in the present study to increase the cut-off point to .50 seconds, in order to give these young lower intermediate-level learners more time for thinking. All pauses were identified again with the help of *NVivo.*

As for the dysfluency component, three dysfluency markers were selected: *repetitions, reformulations* and *incomplete words*. Table 3.11.2 below presents their descriptions, along with the descriptions of the rest of the fluency variables.

Table 3.11.2 Description of fluency variables

TEMPORAL VARIABLES	**Speech Rate**	Total number of L1 and L2 syllables divided by the number of speaking time in seconds.
	Mean Length of Runs	Total number of L1 and L2 syllables divided by the number of runs between pauses in the Presentation part.
	Pause Rate	Percentage of speaking time in seconds spent in both filled and unfilled pauses.
DYSFLUENCY MARKERS	**Repetitions**	Utterances or words produced twice or more without any modification.
	Reformulations	Utterances or words produced twice or more with some modification.
	Incomplete Words	Words that are left incomplete.

As with the WTC measures, intra-rater reliability was sought for the fluency variables, as the personal judgement of the rater was irrelevant to these measures as well. The same sample of students (30 tests), used for the re-analysis of the WTC measures, was used for the fluency measures. Table 3.11.3 below presents the results of the Pearson's correlations conducted for each variable. As with the WTC measures, the first data set was significantly correlated with the second data set, suggesting therefore that the data analysis was reliable.

Table 3.11.3 Correlations for intra-rater reliability: Fluency measures

Fluency measures	Correlations *(r)*
Speech rate	.97**
Mean Length of Runs	.99**
Pause rate	.98**
Repetitions	.97**
Reformulations	.98**
Incomplete words	.97**

**p < .01

ii. *Content variables*

In order to examine the part of the hypothesis predicting that the CS group would attempt more complex ideas, the *content variable* construct was formed. The first variable in this construct was *verb ratio,* which measured the range of verb use by applying the same calculations of TTR (section 3.8.2), but on verb use only. The number of *types* of verbs, that being the different verbs used, was divided by the *tokens* of verbs, that being the total number of all the verbs used. This was multiplied by 100, as a percentage facilitated comparisons. Concerning Richards and Malvern's (1997) criticism on TTR regarding vocabulary exhaustion, this was not expected to become an issue in the present study, as the participants were not asked to produce large amounts of language in the oral production tests. Richards and Malvern (1997) recommended the *D measure* as an alternative to the TTR, but as this is effective only with large amounts of language, it was not selected as a content measure in the present study.

To my knowledge, an exact version of this measure has not been used in the past. The reason I decided to focus on verbs is because I consider a wide range of verb use showing more advanced output, than a wide range of noun use, especially when it comes to lower proficiency learners. For example, a student may produce the following sentence: *I like basketball, football, volleyball, handball and tennis*. Another student may put it this way: *I like basketball. I love football, but I prefer tennis.* If the TTR was employed on all types of words, the score would suggest that the first student had a wider lexical variety, as it cannot show that nouns are added next to each other. A TTR on verbs however, would give a higher score to the second student, who used more verbs.

The second variable in this construct was *L1 syllables.* Similarly to the L2 syllables, this variable was calculated manually by adding up all the L1 syllables produced in each test. This variable was able to determine which group required switching to the L1 more often.

The final variable in this construct was *idea units.* Idea units have been used in the past as a measure of the number of propositions or main ideas in a written or spoken text (e.g. Jacobs *et al.,* 1994; Macaro & Erler, 2008). Table 3.11.4 below shows the three variables of this the content construct, along with their method of analysis.

Table 3.11.4 Description of content variables

Verb Ratio	Percentage of the types of verbs divided by the tokens of verbs in the Presentation part.
L1 syllables	Total number of L1 syllables used in both parts.
Idea Units	Total number of main ideas occurring in the Presentation part.

As with the previous variable constructs, reliability was sought for content variables as well. Intra-rater reliability was sought for verb ratio and L1 syllables, as these were objective variables, but inter-rater reliability was sought for idea units. This is because idea units were judged to be a more subjective variable. Another rater, who was a bilingual speaker of Greek and English, analysed the 30 oral production tests in terms of idea units. The results of the three content variables are presented in Table 3.11.5 below. The high correlations of the three measures suggest a reliable analysis of the data.

Table 3.11.5 Correlations for intra-rater & inter-rater reliability: Content measures

Content measures	**Correlations *(r)***
Verb ratio	.96**
L1 syllables	.98**
Idea units	.95**

**p < .01

iii. *Accuracy measures*

The second part of the research question assumed that the accuracy of the CS group would not deteriorate. Accuracy, which is defined here as 'the ability to produce error-free clauses' (Housen & Kuiken, 2009: 461), has been measured in various ways in the past. One of the common accuracy measures is the *percentage of error-free clauses,* which has been widely used (e.g. Ahmadian, 2011; Foster & Skehan, 1996). It represents the percentage of clauses which do not 'devia[te] from standard norms with respect to syntax, morphology and/or lexicon' (Ahmadian, 2011: 274). Similar to this measure is the *error per T-unit* measure, which was used by Bygate (2001). This measure calculated the incidents of errors per *t*-unit. The higher the number of errors therefore, the less accurate the language was.

Another commonly used measure was the *percentage of correct verb use,* which assumed that verbs were used correctly 'in terms of tense, aspect, modality and subject-verb agreement' (Ahmadian, 2011: 274). Some researchers used the *target-like use of plural* measure (e.g. Crookes, 1989; Wigglesworth, 1997), which was calculated by dividing the number of correctly used plurals with 'the number of obligatory occasions for plurals multiplied by 100'

(Ellis, 2008: 491). Others focused on the use of the *indefinite article* (e.g. Wigglesworth, 1997), which represented the distribution of definite articles in relation to indefinite articles.

For the present study, two of the above variables were selected as the most suitable for this kind of participants: the *percentage of error-free clauses* and the *percentage of correct verb use*. The percentage of error-free clauses was selected because it has been chosen by almost all researchers who measured accuracy in the past, and because it captures most types of errors. The percentage of correct verb use was selected because during the intervention a lot of emphasis was placed on verb use, and more specifically on the past tense. As discussed in section 3.5.1, four out of the eight weeks of the intervention were dedicated to this tense. The two accuracy measures were only used for the analysis of the Presentation part of the oral production tests. This part was considered to give a much more reliable set of data, as the Question/Answer part involved a possible effect of the interviewer's variable.

Intra-rater reliability was again sought for these variables, as the personal judgement of the rater was irrelevant. The results of the Pearson's correlations are presented in the table below. As with the other variables, these are very high correlations, suggesting a reliable data analysis.

Table 3.11.6 Correlations for intra-rater reliability: Accuracy measures

Accuracy measures	**Correlations *(r)***
Percentage of error-free clauses	.96**
Percentage of correct verb use	.98**

**p < .01

Concluding this section, the three variable constructs (i.e. fluency, content, accuracy) are believed to be able to demonstrate not only whether there is a development of fluency, but whether this potential development has any trade-off effects on other aspects of speech production, namely accuracy and the quality of content. The fourth research question assumes therefore that these are aspects of speech production which are interconnected and should be monitored together. This argument will be further discussed in following chapters.

3.11.5. Analysis for Research Question 5

The final research question considered the variation in the L2 proficiency of the participants, in relation to TR+. It asked whether TR+ seemed to benefit high proficiency students more than low proficiency students. The two proficiency groups, which were formed as explained in section 3.8.1, were used for this research question. Having excluded the COM group participants, as they did not repeat tasks or receive any feedback, the twenty five students of the *High Proficiency group (HP group)* who scored the highest on the C-test, and the twenty-five students of the *Low Proficiency group (LP group)* who scored the lowest on the C-test, were selected for the analysis.

The performance of these two groups was analysed in terms of all three variable constructs (fluency, content, accuracy) discussed for research question 4. This would determine whether a potential development of fluency came with detrimental effects on the quality of content or accuracy. The results of the quantitative analysis of this research question are also presented in Chapter 6.

A summary of all the research questions, along with the data collection method and the method of analysis followed for each one, is presented in the table on the next page.

Table 3.11.7 Summary of research questions & method of analysis

Research Questions	Data collection method	Method of analysis
1) What are teachers' views on task implementation in Cyprus, task repetition, metalanguage and the role of proficiency pairs during task completion?	Recordings of teacher interviews	Content analysis of interview data
2) How does the output change from S1 to S2 particularly in terms of content?	Recordings of student interactions during the intervention	*Qualitative analysis:* Comparing the interactions of Session 1 and Session 2 of each week (Analysis related to each condition)
3) a) Does task repetition plus feedback lead to more willingness to communicate than no task repetition? **b) Does task repetition with codeswitching *plus* lead to more willingness to communicate than task repetition with English-only *plus*?**	Recordings of oral production tests	Willingness to communicate - Speaking Time - L2 syllables
4) a) Does task repetition plus feedback lead to more fluency than no task repetition? **b) Does task repetition with codeswitching *plus* lead to more fluency than task repetition with English-only *plus*, with no detrimental effects on accuracy?**	Recordings of oral production tests	*Fluency measures:* - Speech Rate - Mean Length of Runs - Pause rate - Repetitions - Reformulations - Incomplete Words *Content measures:* - Verb Ratio - L1 syllables - Idea units *Accuracy measures:* - Percentage of error-free clauses - Percentage of correct verb use
5) Do higher proficiency students benefit more from task repetition plus feedback than lower proficiency students?	Recordings of oral production tests	Fluency measures *(see above)* Content measures *(see above)* Accuracy measures *(see above)*

3.12. Limitations

There are some limitations regarding the methodology of the present study that should be taken into consideration. First, the absence of random allocation of the participants to the three groups increases the possibilities of having extraneous variables affecting the situation. Although an effort was made to control many of these variables, as well as to make the groups more comparable through the matching process, it is unlikely that all potential confounding variables were accounted for.

Second, the non-probability convenience sampling procedure may potentially compromise generalisability. As Dörnyei (2007) puts it, 'the extent of generalisability in this type of sample is often negligible' (99). Nevertheless, all three schools participating in the study fit the criteria of the target population; they are all Greek-medium state primary schools in urban Nicosia. It could be argued thus that since the schools match the population, the results can be generalised across this population.

Third, the relatively small sample size makes it less likely for the quantitative results to yield statistical significance. Therefore, the possibility of this involving a Type II error should not be ruled out. It could be the case that, due to the size of the sample, the statistical tests might fail to detect a difference in the population, when in fact there is one (Field, 2009).

In addition, although an effort was made for the teacher variable to be controlled for, there is a possibility that this could still affect the situation. All the lessons were observed and some of them were recorded. Although the recordings offer some indication that the linguistic behaviour of the two teachers was very similar, their compatibility could still be questioned, as not all lessons were recorded. What would strengthen my argument would be recording, or even video-recording, all the lessons throughout the intervention.

Moreover, the *school effect* should be considered. When different schools act as different groups in an intervention, there is always the possibility of the school itself becoming a confounding variable. Although there are no performance tables of state schools, one school may be of a higher standard than the other, in terms of the teachers' qualifications or experience, the strictness on students' behaviour, the parents' involvement and so on. These are variables that could not be controlled in the present study, but that should nevertheless be considered.

Furthermore, my presence in the intervention classrooms might have possibly affected the situation. In order to minimise this effect, I had paid regular visits to the three schools before the intervention started. During this period of time, I tried to approach the students that would participate in my study, as well as the teachers, and give them the opportunity to get to know me. Despite these efforts however, the fact that a new person was present in their classroom could have affected the behaviour and performance of the students during the lessons, and consequently during the completion of the tasks. This possible effect should be kept into consideration.

Finally, there is always the issue of the *researcher's bias* in research. As the sole researcher of this study, I did all I could to remain subjective towards all three groups, despite the fact that based on my hypothesis I had certain expectations. I created and followed the researcher's protocol discussed in section 3.9.1, in order to make sure that the same steps were taken in every session of every group throughout the intervention. This ensured that one group was not favoured over another. In addition, when analysing the data, I tried to remain as objective as possible. Nevertheless, researchers may be subconsciously biased towards the group that could prove their hypothesis correct. Therefore, this is an unavoidable limitation.

3.13. Chapter summary

The chapter began by formulating five research questions for the present inquiry, based on the discussion of the literature and the hypotheses of the study. These research questions aimed to introduce the context of the study in much more depth by examining the teachers' perceptions on task implementation and other related issues. They prompted a qualitative investigation into the students' interactions when the task is repeated with feedback. In addition, they inquired about the effects of the three independent variables, namely "task repetition plus", "task repetition with CS plus" and "task repetition with EO plus", on WTC and on the development of L2 oral fluency. Finally, the question of whether TR+ benefits more high proficiency or low proficiency students was also under investigation.

The chapter continued with a description of the research design; a mixed methods approach employing a quasi-experimental design and a qualitative analysis of task language and interview data was adopted. A detailed account of the intervention was given, with reference to the three conditions, as well as the oral production tests used as pre- and post-tests. A discussion on the measures taken for fidelity to conditions was included. Moreover, a description of the pilot study justified some of the decisions taken for the main study, in terms of the task types, the oral tests and the conditions. The procedures followed for recruiting a matching sample to the target population were explained, as well as the ethical procedures that were considered.

As this was a classroom-based study, an effort was made to match the three groups in terms of their proficiency, their productive vocabulary, their socio-economic status and their prior contact with the English language. Other possible confounding variables that could affect the intervention were also considered. In complementation to the main study, the procedures followed for the collection of interview data from English teachers in Nicosia were described.

Following the data collection part, the data analysis procedure that was followed for each question was discussed. The data for the first two research questions were analysed qualitatively. This approach was considered to be more suitable in detecting the themes emerging from the teacher interview data, as well as in investigating whether TR+ would offer the opportunity to the participants to improve the content of their output. For the last three research questions, a quantitative analysis was conducted. The variables selected for the analysis were described, while reference was made to softwares that aided the analysis. The chapter ended by referring to the limitations of the study.

CHAPTER 4: Qualitative Analysis of Teacher Interviews

4.1. Introduction

This chapter provides a response to the first research question, which inquires about the teachers' opinions on issues emerging from the literature review. As discussed in the Methodology chapter, nineteen teachers, who were teaching English to sixth grade students in urban schools in Nicosia at the time of the study, were interviewed. More specifically, the teachers were asked about four main topics: task implementation in Cyprus, task repetition, metalanguage and the role of proficiency when students work in pairs. As the interviews were semi-structured, the teachers were free to refer to any other topic that they felt was relevant to our discussion.

The chapter is divided into four sections which represent the four main topics of discussion. Having conducted the interviews in Greek, certain parts of the interviews were transcribed and translated to English by myself. Excerpts from the transcripts are used throughout the chapter to illustrate the opinions of the teachers. All names that appear in the chapter are pseudonyms, in order to protect the identity of the participants. By qualitatively analysing the teacher interview data, the chapter provides a more profound understanding of the context of the study.

4.2. Task implementation

4.2.1. The value of tasks

At the start of the interview, I aimed to gather information on the general opinion of teachers on task implementation within the Cypriot primary education context. More specifically, I attempted to understand whether any kind of tasks were already being used and what the

teachers thought the value of those tasks was. In order to ensure that with the term "tasks" the teachers and I referred to the same type of activity, I gave them my definition of tasks at the beginning of the interview. In particular, I described tasks as oral activities completed in pairs or small groups, which required students to use the target language creatively in order to reach a pre-defined outcome.

Having reviewed all recordings, it seems that all the teachers were in favour of tasks for many reasons. From a more practical perspective, teachers found that tasks interest students because of their more playful form. Students, especially at this age, seem to enjoy working in pairs or small groups much more than completing exercises on their own. Despite the practical benefits of tasks however, teachers also acknowledged a pedagogical value in the use of tasks. It emerged through the interviews that teachers prioritise the development of oral production and communication skills in the target language, as it is recommended by the teacher's book, and the use of tasks is one way of achieving this goal.

The acknowledgement of practical and pedagogical benefits of tasks is well represented in Ellie's interview. Ellie stated that tasks were perhaps the only activities that students were interested in because of their more playful form, rather than the activities of the book. She continued by giving the example of activities that required students to guess or discover something. She characterised the students' interest in working in pairs or groups as "completely natural" because, in her view, language acquisition was dependent on communication. According to her, language could not be learnt without communication. Tasks therefore should not just be present in the classroom, but they should be used a lot more. Her observation on students in this context was that although they might know the target language well, that is knowing how to apply rules and possessing a range of vocabulary, they get stuck when it comes to speaking due to embarrassment. She

characterised this as the “eternal problem” with students in this context. In the following quote therefore, Ellie explains why she places more emphasis on oral practice.

Excerpt 4.2.1. Ellie’s opinion on tasks

> *So I would personally like it, if they had more opportunities to interact; to say their opinion, or to say what they think [...]. I place a lot of emphasis on the oral practice of the language [...]. Because I find that it will be more useful to them. We don’t know how many students will need to use the language in written form in the future. Oral practice however is more important because it concerns everyone [...]. I believe that everyone should at least be able to speak and communicate [in English].*

Overall, having recognised that students enjoy tasks more than any other activity in the lesson, Ellie admitted that she would have liked to implement more such activities in her lessons. By completing more tasks, students could overcome feelings of embarrassment, which often occur when they speak in the target language. She acknowledged the importance of practising oral communication when learning a foreign language, and believed that this was the aspect of a language that all students would need the most in the future.

Another example of the benefits that teachers can find in the use of tasks comes from Anne’s interview. Anne viewed tasks as a good opportunity for students to practise what they had learned in the lesson within a freer communicative context. She reported that her students were also enthusiastic about these sorts of activities, which they found very interesting and motivating. The excerpt below illustrates her opinion.

Excerpt 4.2.2 Anne’s opinion on tasks

> *Of course I think that it is a very important activity in the English lesson. For this context to exist...the freer context; the communicative. So [the students] can use the structures they learned in the lesson within a freer context of communication. This makes the lesson more interesting. Because [tasks] have this more playful form, they provide a motive to the children to use it; they involve the children with more enthusiasm in the activity.*

Another teacher, Mona, referred to the cognitive challenges that tasks provided to her students. According to her, tasks offered the opportunity to students to think about which L2 words were appropriate in different situations. Mona also referred to the psychological benefits that students could have from completing tasks. More specifically, she argued that the playful form of tasks made them seem easier and more attractive. Students therefore were much more likely to make an effort to complete them and overcame their feelings of embarrassment. Her opinion is presented in the excerpt below.

Excerpt 4.2.3 Mona's opinion on tasks

> *Yes [the students] are very much benefitted [from tasks] in many ways. [They are benefitted] in English, in the sense that their brain works in order to find words...which word fits in each situation; but it is also good for their confidence because some students are ashamed of coming in front of the class to say something. When it has a more playful form, it is easier for them to come in front of the class.*

In general therefore, it seems that teachers in this context find some kind of value in the use of tasks for a number of reasons. Tasks are not only very interesting and enjoyable to students, but they are also helpful in terms of the oral development of the target language, which is a very important skill to these teachers. In addition, the teachers believe that tasks help boost the confidence of their students, who are more likely to overcome feelings of embarrassment when they initially communicate in this language with their peers. Moreover, tasks offer a cognitive challenge to students, who try to access the appropriate L2 vocabulary for each situation.

4.2.2. Factors affecting task implementation

Having established that teachers acknowledge the importance of L2 oral practice and tasks as a way towards this, teachers were subsequently asked about the extent to which tasks are being used in their classrooms. The majority of the teachers responded that working in pairs or groups is a very frequent activity in their lessons because students enjoy them very much. However, through our discussions, it emerged that a number of factors affect the situation of task implementation.

First, the teachers referred to the general inadequacy of the current books. Most of the teachers characterised them as "dated" with material that does not interest students of this age. In terms of tasks specifically, the teachers noted that the oral activities contained in the book do not provide opportunities for creative oral communication. Instead they guide students' output to a large extent, allowing thus for very limited creative speech to take place. For example, Mona stated that most of the tasks in the book were guided. Although they asked students to work in pairs or groups, they did not allow freedom of expression. The opinions of two more teachers, presented in the following excerpt, illustrate the issue of the books.

Excerpts 4.2.4 Teachers' opinions on tasks in the book

> Laura: *"Activities that are very creative and that will recap previous knowledge, build on it, as well as use new knowledge are very limited. Most of them [in the book] are much guided and are done mechanically [...]. There are not many."*
>
> Vanessa: *"The way [tasks] are in the book is very much guided. They lose a lot from their creativity and they end up very tiring [...]. There is no flexibility on vocabulary. Essentially the activities are very typical. And what is also very negative is that exactly the same tasks are repeated in the workbook. There is no differentiation in the workbook. So the teacher should do something of his/her own. Or else there will be no consolidation."*

From the excerpts above, it is clear that teachers find tasks in the book very guided, as they allow for no creative language and free speech to take place. In other words, they provide students with what should be said. It would be fair to say therefore, that the activities contained in the book can hardly be characterised as communicative; much less as tasks. Although the term ‘tasks’ is used throughout the chapter, it refers to oral activities completed in pairs that are not largely communicative.

In an attempt to explain why the tasks provided by the book are so guided, another issue emerged; that of the huge gap of L2 proficiency among the students. According to some teachers, the reason tasks lack open-endedness is because lower proficiency (LP) students should be able to take part in all activities going on in the classroom. The more open-ended a task is, the harder it becomes for LP students to contribute to its completion. The existence of mixed-proficiency classrooms in public schools is due to the fact that the majority of the students, who are more advantaged students, attend private lessons in English after school. These students often start their private lessons two or even three years before English is introduced at school. As Jane stated, some students might even have had a maid at home with whom they communicated in English. At the same time, there are less privileged students who did not have the opportunity to attend private lessons, or who did so much later. As Jane said, these students might not have been encouraged by a member of their family to learn a foreign language.

It is not surprising therefore that some students end up with much more knowledge of the L2 than their peers at fourth grade in primary schools. As Anne stated in her interview, the issue of different proficiency levels was very obvious and problematic. The teacher was always bound to “lose” certain students at some point during the lesson: either the more proficient students, who found what they did very easy and lost interest in it; or the less proficient students who could not follow when what they did was more advanced. However, as many

teachers noted in their interviews, schools must follow the knowledge of the average student and this is why the books contain tasks that restrict the creativity that higher proficiency students could have produced. The following quotes coming from the interviews of two different teachers present well the issue of mixed-proficiency classrooms.

Excerpts 4.2.5 Teachers' perceptions on the issue of mixed-proficiency classrooms

> Laura: *"The reason they are more guided and more specific is because we have to follow [the average student]. We don't start with those who already know a lot and build on them. We start with the average student. We want the average students to understand the language and the weaker students to gain something. So the level is not the one that it should be for 5th or 6th grade, because most of the children....we use in school the knowledge that we haven't taught them. Most of the children go to private lessons; they start for example at 2nd grade. English in public schools starts at 4th grade [....]. These activities take place but they are limited because the time is limited. We are talking about two periods per week. Forty minutes [each] and we have the different [proficiency] levels. So [doing tasks] is difficult because when there are many weak students, then the time you'll dedicate to such creative activities is very limited because creative tasks demand a good foundation of knowledge."*

> Mona: *"I imagine that the reason they do this is for weaker students to be able to participate because this is a [public] school. It is not like private lessons, where they might move faster or give them some kind of [additional] work. We also do that but all students should participate [...]. [Tasks] are very much guided. [The book] gives the words, the pictures and the structure of the question or the answer [...]. They are not particularly creative."*

Apart from the issues of inadequate books and of mixed-proficiency classrooms, it is worth noting the issue of limited time dedicated to the subject of English; an issue raised in Laura's interview shown in the excerpt above. Laura says that the lessons of English take place only twice a week and that each lesson lasts for forty minutes only. Due to this limited time, it is difficult to do tasks, which require time, especially with mixed-proficiency classrooms. This is probably because tasks require more time to explain what has to be done, than simple individual activities found in the book, due to their more complicated structure. Students need to understand what the outcome of the task should be before stating to work on it. Moreover,

the fact that tasks allow students to negotiate about what has to be done, could often be time-consuming. Another thing that was mentioned in a few interviews and could be related to the time requirements of tasks is the possibility of students getting off track and start conversing about unrelated issues when working in pairs. This could again be one of the reasons why tasks need more time to be completed.

Despite the students' many years of private lessons in English, two teachers, Ellie and Vanessa, noted that students were not particularly good at communicating orally in the L2. According to them, the reason was because the lessons they attended were exam-oriented, as requested by the parents. What the parents want is for their children to obtain certificates in English, which can be used in the future as proof of acquisition of the language. According to the two teachers therefore, students become very good at completing exercises, but lack communicative skills.

As Ellie said, although students start private lessons from grade 3, if not from grade 2, very few speak the language with the fluency that corresponds to the years that they take private lessons. She continued by characterising this situation as "weird" due to the extended contact students have with the English language. The excerpt below presents her opinion and it is followed by that of Vanessa, who also referred to this problem.

Excerpts 4.2.6 Teachers' perceptions on students' lack of communicative ability.

> Ellie: *"This is weird because first of all they come to contact with the [English] language very often because of films, tv series, in which almost everything is in English. They come to contact through the electronic games they play. And also through the music they listen to. So reasonably it should have been very easy for them to express themselves in English."*
>
> Vanessa: *"I think speaking is the most important. Our students go to private lessons for many years but they cannot compose correct sentences. It is tragic [...]. I think [private educators] only prepare the students for exams. This is what the parents ask for. The children learn how to solve exercises in exams. This is very clear from the listening we do in class. They might not understand exactly what is being said, but*

> *they can complete the tables. They are very well-trained on certain things. But they cannot speak, nor structure a sentence."*

As shown in the excerpts above, Ellie and Vanessa believed that their students should have been more skilled in communicating in the L2 orally, given the many years of private lessons they attend and the exposure they have to the English language in their everyday lives. They attributed this lack of communicative skills to the fact that private lessons had become exam-oriented and taught techniques about how to complete exercises, rather than developing the language itself. Nevertheless, the approach adopted in private language schools should not be unexpected because, as mentioned by a few teachers in their interviews, parents demand that their children successfully pass exams. It could be said therefore that parents seem to be less concerned about whether or not their children learn the language profoundly in order to be able to communicate in that language. It is not surprising therefore that private language schools devote time into practising exam activities, rather than oral communication activities and that the students' oral communication skills are inferior to their written skills.

4.2.3. Experiences of teaching

Having reviewed all interviews, I understood that since the book does not provide opportunities for creative L2 speech, the extent of creativity of the tasks given to the students largely depends on teachers. As mentioned in a few interviews, teachers are allowed by the inspectors to go beyond the book and bring their own material if they wish to, as long as the curriculum is covered. In other words, teachers have the freedom to bring or create their own material depending on the proficiency level of their classroom and the extent of creative language they want their students to attempt to produce.

The discussions of this lack of adequate material and of the freedom of teachers to create their own brought to the surface a number of other related issues. To start with, the teachers referred to the fact that they teach a variety of subjects in primary schools, for which they do not have enough time to prepare as much as they would have liked to. These teachers might teach anything from mathematics to music. So for some subjects, they prepare less and usually these subjects are the ones conducted once or twice a week, such as English. English is viewed as one of the secondary subjects in primary education, so teachers in general do not dedicate a lot of time preparing for it. An excerpt from Vanessa's interview demonstrates this problem well.

Excerpt 4.2.7 Teacher's view of English as a secondary subject

> *Vanessa: "I, for example, teach a lot of subjects because I am young. I cannot prepare perfectly for all subjects. So I'll "cheat" on subjects of this nature, i.e. English, music, art. So you realise that you can't be perfect for subjects like mathematics and history, as well as secondary subjects [...]. If I had the time, if I were an English teacher, I would have found a lot of material for English. But unfortunately I am not just an English teacher and for me English, in relation to the other subjects I teach, comes second."*

Although primary schools teachers are expected to be able to teach all subjects at all six grades of primary education, some teachers mentioned in their interviews that they do not believe that such a "multi-competent" teacher exists. Having received no formal training in teaching foreign languages at university or elsewhere, many admitted feeling the lack of skills to teach a foreign language. Most of them even mentioned that they were "forced" in a way to teach English by their colleagues. When no other teacher wanted to teach the subject, they were chosen as more suitable in their schools, either because they had earned a certificate in English (e.g. GSE O'level) many years ago, or because they had spent a year in an English-speaking country studying for a Master's degree often in another subject. Having gone through the experience of teaching a foreign language, some of the teachers mentioned that they do not feel competent enough to teach it. This feeling could become more intense when it

comes to task-based language teaching, which is a relatively newer approach to language teaching that requires time for preparation and experience.

Excerpts 4.2.8 Teachers' perceptions on their lack of expertise

> Victoria: *"At university, we did not have an ELT course. So we teach based on what we do at training and based on what we hear from colleagues who went to some other training. So I would say that [training] is important. Greek is learned in a different way than English."*
>
> Jane: *"[Training is] definitely [required]. The first time I taught English, I was given the subject because no-one else would take it and I had a GCE. The GSE is not a qualification to teach English. The fact that I obtained a certificate doesn't mean that I feel more familiar and comfortable with the language. I can communicate with a native English speaker, I can write a text [in English], but to conduct a lesson is a big deal [...]. There are students who know better English than we do."*

As shown in the excerpts above, teachers admit not feeling competent enough to teach English, as they are not adequately qualified. Jane even commented at the end of the excerpt that there are students who know better English than their teachers. Due to these feelings of their lack of expertise, Stephanie for example, showed her disagreement with the concept of the "multi-competent" teacher, as she did not believe that it was possible for a teacher to be able to teach all subjects at all levels. She further commented that there should be some kind of specialisation of teachers to certain subjects.

These are problems of the educational system in Cyprus that arose throughout the teacher interviews. Overall however, these issues relate to task-based learning in two ways: first, because teachers teach a variety of subjects, they do not dedicate time into finding appropriate creative tasks for their classrooms; second, many teachers do not feel competent enough teaching a foreign language, let alone adopting a task-based language teaching approach.

Since the creativity of tasks depends on teachers and since teachers lack time and expertise, many mentioned that they would have liked it if a source of materials and activities was

available to them by the Ministry of Education in Cyprus, from which they could select the ones that are appropriate for their own classes. This request for more available material also emerged when the teachers were asked about whether any additional training is needed for tasks. More specifically, the majority of the teachers replied that what they need is more ideas for creative activities. Whether these come from a source of materials or teacher training courses, the teachers' unanimous answer was that they need help with coming up with new ideas.

Excerpts 4.2.9 Teachers' request for more creative ideas

> Mona: *"If [teachers] want to do something extra, in the sense of creative tasks that you talk about, then [training] is required. [We need] some ideas."*
>
> John: *"I find that when I attend a seminar or a conference or when I watch a sample lesson [...], [I find] something that I can apply to the students I have, at the level I have. It definitely helps to change, to listen to others. And we teachers get bored doing the same things every year. We want a change as well [...]. So education is always necessary for [the teachers] to think of something new, something more creative, something that would fit with the level of their learners."*

A few teachers however, did not believe that they needed additional training courses. These teachers felt that the teachers' book was enough for conducting a lesson in English and that additional training was not really necessary. According to them, whenever they faced a difficulty or whenever they were unsure of something, they consulted the teachers' book.

4.2.4. The new curriculum

As mentioned in the introductory chapter, a new curriculum was implemented in September 2012; a few months after the data collection period of the present study. Many teachers referred to this upcoming new curriculum in their interviews, as tasks seem to be central in its implementation. According to the information I gathered from the teacher interviews,

English was to be introduced at pre-school in September 2012, rather than 4th grade, which was the case at the time of the study. This introduction took place in the oral and aural form of the language only, through story-telling and songs in English. The students will continue English at grade 1 of primary school and then throughout their primary school years. The new curriculum also features new books for all grades of primary school, again prepared by the Ministry of Education and Culture of Cyprus.

With this early introduction to the language, most of the teachers expressed their hope that until grades 5 and 6, students would feel more confident and more able to express themselves in English. At that point, more creative tasks are more likely to be implemented effectively.

4.2.5. Conclusions on task implementation

Concluding this section, it seems that teachers of primary schools in Cyprus believe that tasks are very important because of the opportunity they provide for practising oral communication in the target language. According to the teachers, apart from the practical benefits of tasks, which involve stimulating the students' interest, such activities provide psychological benefits, as well as cognitive challenges to the students. For these reasons, most of the teachers use oral communication activities in their classrooms, which involve pair or group work. Nevertheless, none of the teachers mentioned whether they expected the benefit of oral production to be much clearer in specific aspects of speech production, such as fluency or accuracy.

When comparing these results with those of past research, mainly the studies conducted by Carless and colleagues (e.g. Carless & Gordon, 1997; Carless, 2002, 2003) in the context of

primary schools in Hong Kong, it seems that the results are partly compatible. While teachers in Honk Kong agree with the teachers in Cyprus in that students find tasks particularly interesting and enjoyable, they do not think that tasks provide additional learning opportunities than any other activities. On the contrary, teachers interviewed in the present study, stressed that tasks are essential for the development of their students' oral communication skills. It seems therefore that teachers in these two contexts disagree in terms of the pedagogical value of tasks.

Furthermore, being asked whether tasks were used in their classrooms at the time of the study, the teachers raised a number of issues that are relevant to task implementation. They referred to the issue of mixed-proficiency classrooms, which constitutes a real problem in language education in primary schools. As a large number of students are introduced to the English language through private lessons from a very young age, a huge gap of proficiency level features in the classrooms. Teachers therefore must follow the knowledge of the average student in their lessons.

In order for students of all proficiency levels to be able to participate in everything that is going on in the classrooms, communicative activities featured in the books are very much guided and do not allow for much creative speech to take place. If teachers wish to conduct more open-ended activities in their lessons, then they are the ones responsible for finding and preparing adequate material. Being aware of this responsibility, teachers referred to their lack of time to prepare adequately for all lessons, to their lack of expertise to teach a foreign language and to their need for more innovative ideas. Some of these issues were also mentioned in Carless' (2002) study by the teachers in Hong Kong. They referred to tasks being time consuming both in their preparation and implementation.

It is clear therefore that there are a number of issues involved to the question of whether tasks are being implemented or can be implemented in this context. Teachers mentioned however that with the new curriculum, it would be easier for them to implement tasks in the future.

4.3. Task repetition

The second category of questions the teachers were asked, concerned task repetition. More specifically, teachers were asked whether it is common for them to repeat tasks in their lessons and if they thought that their students could gain something from task repetition. Almost all teachers replied that they usually repeat the same type of tasks, but on a different topic; a different conceptualisation to that of Bygate's (2001) concept. The reason they choose not to repeat exactly the same tasks is because they find that students lose interest in what they do.

One of the teachers who shared this opinion was Stephanie. Stephanie stated that completing the same task on exactly the same topic could be boring. She further commented however, that repeating the same types of tasks on new topics helped students become familiar with the types of tasks. On subsequent occasions therefore, the problem of getting to know the task was overcome and students were more able to focus on the language. The opinion of two more teachers is presented in the excerpts below.

Excerpt 4.3.1 Teachers' overall perceptions on task repetition

> Vanessa: *"If it's just similar, then it could be beneficial. If it's the same, the children tend to remember very much. And often they are bored doing the same things. They feel that since we already did this, it's over. If it's similar however, with different adjectives, different verbs, they can do it."*
>
> Laura: *"I think there is a benefit [when repeating tasks] because they refresh their memory regarding the structures we worked on during the last lesson [...]. And this helps weaker students as well because things that they might*

> *have not comprehended well on the last lesson are restored. They fill in gaps. And the strong students practise even more [...]. We don't use the same vocabulary but the structure is the same. We simply change the vocabulary. [The students] build on previous knowledge [...]. You can't do the same activity with the same vocabulary. There is no point [...]. [Repeating tasks] makes them feel satisfied because it is on something they've already done [...] and they are able to complete it with new vocabulary."*

According to Laura's opinion therefore, students of all proficiency levels had something to gain from repeating the same type of tasks. Lower proficiency students had the opportunity to fill in potential gaps by refreshing structures that they had not comprehended well on previous sessions, while higher proficiency students had the chance to practise their oral skills even more. This positive attitude towards repeating the same types of tasks was held by almost all the teachers interviewed.

While sharing the opinion of her colleagues on same task type but different content, another teacher, Alison, clarified that she only chose to repeat exactly the same task when it was a game that was completed at the end of the first lesson on a unit. This could be repeated at the beginning of the second lesson as an adequate continuation of the previous lesson. However, Alison pointed out that she did not expect this repetition to fulfil an important goal of the lesson, and that it could not help higher proficiency students in any way. She believed that it could only have psychological benefits for less proficient students. Recalling the correct answers helped boost the confidence of these students, and made them start feeling more comfortable with the target language. Her opinion is illustrated in the next excerpt.

Excerpt 4.3.2 Alison's view on task repetition

> *They do the same type of task many times but not with the same goal. The grammar use or the tense differs [...]. I repeat exactly the same task when it is a game and it is completed at the end of a lesson. I might start the following lesson with the same game as feedback [...]. The weak students are benefitted [...]. When they remember things [they are benefitted] for psychological reasons. It boosts their confidence [...] but I cannot tell you that I expect this to fulfil an important goal. I simply think that [...] the strong students who remember the answers do not benefit in any way but the weak students who remember the answer will learn something.*

Other teachers also referred to occasions where they repeated exactly the same task that the first lesson on a certain unit ended with, as the beginning of the second lesson on that unit. One of the teachers, Rebecca, found that it was easier for her students to complete the task on the second occasion because they had already made an attempt. She also added that when her students enjoyed a task, they asked to repeat it themselves. Another teacher, Catherine mentioned that tasks were often repeated at the beginning of the following lesson because she thought that repetition was good for reminding students what they had already done before starting something new. The excerpts from their interviews referring to these comments follow.

Excerpt 4.3.3 Rebecca's & Catherine's view on task repetition

> Rebecca: *"Sometimes yes. I will not repeat something after five lessons [...]. If a task is completed at the end of a lesson, then on the following lesson I will start with that and I see that it is easier for the children to do it. An attempt has already been made, it might not have been perfect; but the second time I find that it is easier for more students to complete it [...]. If it's something they really enjoyed, they ask for it themselves."*
>
> Catherine: *"We usually do [repeat tasks]. On the first lesson, you teach a new phenomenon with some activities. The next lesson begins [...] with the previous lesson in order to go into something new. It is good to do some kind of repetition to remind the students what they previously did."*

Concluding on the task repetition part therefore, it seems that in general teachers in this context prefer to repeat the same types of tasks on different topics, rather than exactly the same tasks. The reason most of the teachers rejected same-task repetition is because of the loss of interest on the part of the students when they do not work on something new. The reported benefit of repeating the same type of task on a different topic was the fact that it offered the chance to focus on the language more, once task type was familiar.

A few teachers however, referred to certain occasions where exactly the same tasks are repeated. These include cases where a task was completed at the end of a lesson for the first

time and then the following lesson started with the repetition of that task as a continuation of the previous lesson. Teachers considered this to be a better introduction to the new lesson. Nevertheless, as mentioned by one teacher, the only benefits that could be achieved from repeating the same tasks were psychological; more specifically, boosting the confidence of lower proficiency students. No major goal regarding the language learning process was expected to be achieved.

It is worth noting that none of the teachers referred specifically to potential oral production gains from task repetition, namely gains in terms of fluency, accuracy or complexity. Bygate’s (2001) argument indicated that the minimum that should be expected to happen from task repetition is the development of fluency. It seems that the teachers in the present study therefore, had not noticed any such gains.

4.4. Metalanguage

The third major topic of the teacher interviews concerned metalanguage and the extent to which this occurs in the target language. All teachers agreed that this kind of language always, or most of the time, occurs in their mother tongue. When students communicate between them, they spontaneously do this in their mother tongue.

Some teachers added that it is not feasible for metalanguage to occur in the target language. In an attempt to explain why this is, some mentioned that, while students already know how to give simple instructions in English, they need time to structure their sentences because they are not used to using this kind of language in the L2. When they have to complete a task within a given time therefore, they cannot spend most of it trying to communicate about what

they have to do. So the issue of limited lesson time mentioned earlier as a barrier for task implementation by the teachers could be linked here.

Another reported reason why metalanguage always occurs in the L1 is the gap of proficiency level between students. Acknowledging that using L2 metalanguage requires a certain level of proficiency, teachers commented that when the interlocutor of a student who uses L2 metalanguage is a less proficient student, it is likely that the latter will not understand the L2 instructions or questions. The conversation therefore would then need to be switched to the L1 for effective communication.

Overall, students tend to switch to their mother tongue for metalanguage because it is faster, given that tasks are completed within time limits, and much more efficient in mixed-proficiency groups. The following excerpt comes from the interview conducted with Rebecca, who summarises these two reasons.

Excerpt 4.4.1 From Rebecca's interview

> *But they already know [how to use metalanguage in English]. It's not difficult to say "You have to say 'this is book'" for example. They can give simple instructions in English. But when I tell them 'You have three minutes to prepare something' and they try to prepare it in three minutes, it's more time-consuming to think of the instruction in English and use it. And if their interlocutor is a weaker student and does not understand it, they'll say it in Greek. I don't find [using L2 metalanguage] impossible. I find it unlikely.*

Another teacher, Jessica, shared Rebecca's opinion on time constraints and proficiency level. She added however that using L2 for metalanguage also depended on the relationship that the students have in a classroom. When students feel comfortable with each other, then communicating in the L2 for metalanguage is more feasible. The excerpt where she mentions this follows.

Excerpt 4.4.2 From Jessica's interview

> *It depends on the students and on the classroom. When they feel comfortable with each other, it's more feasible. But I think most of them are not ready for that [...]. There are students who get "lost" with English instructions and stop trying. But between them it is very hard, because they don't have the same proficiency level. So if you ask them to try to communicate between them in English, they will spend most of the time trying to communicate in English, rather than completing the task.*

Throughout the interviews, teachers mentioned at several points that students of this age, which is pre-puberty, get embarrassed very easily and make fun of their peers very easily. This is why Jessica believed that a good relationship between students was also crucial for effective communication in the L2.

While most teachers, including the ones cited above, seemed to think that using L2 metalanguage was very unlikely to happen in this context, Anne was more optimistic about it since she believed that this was something that depended on the teacher. The following excerpt presents her opinion.

Excerpt 4.4.3 From Anne's interview

> *I think that this starts from the teacher. When [the teacher] uses English exclusively and all the instructions are given in English and s/he insists [on English], I think that it can be achieved to a certain extent. Even when [the students] ask him/her something in Greek, s/he insists to say that in English [...], [it's more feasible]. Some [of these phrases] could even be written on the board or on cards.*

According to Anne therefore, students' use of L2 metalanguage is something that depends on the teachers' practice. When teachers use English exclusively when they give instructions and when they insist that their students speak in English and ask questions in English, then this is more likely to happen. However, Jane happened to mention that she had tried this method before and in her case, it was not very effective. The following excerpt refers to her experience.

Excerpt 4.4.5 From Jane's interview

> *I tried to do this, this year; to enter the classroom and speak in English only and not in Greek. The first month I did this constantly; I insisted. Then I realised that the only thing that happened is that they were making fun of me. The ones who felt uncomfortable felt even more uncomfortable. They expected from the others to tell them in Greek what was going on.*

As shown from the excerpt, her effort to stick to English herself while teaching the language was not effective. Lower proficiency students who felt uncomfortable with English, felt even more uncomfortable when their teacher never switched to Greek. Rather than trying to understand however, they waited for their higher proficiency peers to explain to them in Greek what was going on. When Jane realised that her method was not working, she decided to abandon her efforts and allow herself to switch to Greek when necessary. It seems therefore that Anne's suggestion regarding the teachers' exclusive L2 practice is not always feasible in this context.

Finally, some teachers again placed their hopes on the new curriculum that would start being implemented the following year. As this curriculum introduces English to students from a very young age, teachers expected that by the age of twelve, students would be feeling a lot more comfortable with speaking the language among them. At that point, they would be in a better position to communicate in the L2 for metalanguage.

One of the teachers referring to the expectations she had from the new curriculum was Stephanie. According to her, the students would start familiarising themselves with simple L2 vocabulary (e.g. numbers, animals) from Grade 1. This would come without any pressure because, as she said, the students in first and second grade still try to acquire the mechanisms of producing mainly written, as well as oral, language in Greek. Stephanie stated that this pressure-less familiarisation with the L2 would help students feel more comfortable with speaking the L2 when they reach higher grades of primary education. Under such

circumstances, metalanguage was more likely to occur in the L2. Stephanie's view was also shared by Jane, whose opinion is presented in the excerpt below.

Excerpt 4.4.6 Teachers' expectations on the new curriculum

> Jane: *"[Using the L2 for metalanguage] might be successful from next year when students start learning English from pre-primary school with story-telling. It will not be teaching; it will be a familiarisation with the language in the initial years. In this case, where all the children will start English together from a young age, I think this goal of metalanguage will be achieved."*

Overall, most of the teachers believed that the new curriculum would prepare the students better for L2 oral communication once they reach sixth grade. The students would start familiarising themselves with the language from a very young age through singing and story-telling in the L2. This is expected to help them become much more comfortable with speaking the language.

In conclusion, according to these teachers, students' use of metalanguage always occurs in Greek. While in the interviews none of the teachers acknowledged a role for L1 metalanguage in the execution of tasks, most of the teachers seemed to be understanding, by referring to the reasons why this occurs in the L1. Two main reasons they reported were the limited time students were given to complete task, which did not allow them to spend time working on their L2 metalanguage; and the mixed-proficiency level, which minimises the possibilities for remaining in the L2. Linking these results with the views of teachers in Hong Kong, it seems that, although both groups of teachers do not refer to the potential facilitative role of L1 metalanguage, they disagree in terms of the reasons why this occurs in the L1. The teachers in Hong Kong view their students' L1 use as their own 'failure' (Carless & Gordon, 1997), while teachers in Cyprus refer to time limits and mixed-proficiency as reasons why this occurs in the L1. Finally, concerning the teachers' views on the potential of metalanguage being expressed in the L2, most of the teachers referred to the new curriculum, as the way towards the L2 being used for metalanguage.

4.5. Combinations of pairs in terms of proficiency

4.5.1 Teachers' views on mixed-proficiency and same-proficiency pairs

The last category of questions in the teachers' interviews concerned the combination of pairs, in terms of proficiency, that the teachers believed were most effective. Given the variety of proficiency levels in these classrooms, teachers were asked about whether they think that same-proficiency pairs worked better than mixed-proficiency pairs. The majority of the teachers replied that they preferred mixed-proficiency pairs because in this way higher proficiency (HP) students could help their lower proficiency (LP) peers. The excerpts below illustrate their opinion.

Excerpts 4.5.1 Teachers' preference for mixed-proficiency pairs

> Nichole: *"When a same-proficiency pair consists of high proficiency students, the result is perfect. However, I try to have mixed-proficiency pairs in order to make lower proficiency students practise and have some kind of corrections from the more able students [...]. I prefer mixed-proficiency pairs. I find that it works better."*
>
> Victoria: *"Usually you must have a good student with a weaker student because very often when two weaker students work together, they converse in Greek about things other than the task at hand [...]. Definitely mixed-proficiency pairs [...]. They collaborate and the stronger student helps the weaker student. The weaker student can learn with the help of the stronger student."*

As shown in the excerpts above, teachers believe in *scaffolding* (Antón & DiCamilla, 1999) and prefer mixed-proficiency pairs in order to facilitate it. In terms of same-proficiency pairs, the teachers recognise the benefits of having two HP students working together, but they can also see the disadvantages of having two LP students working together. In such cases, it is much more likely for task language to be replaced with off-task talk in the L1.

Although most teachers expressed their preference for mixed-proficiency pairs, some of them referred to other factors that they considered as pre-requisites for successful scaffolded help

to take place. The first factor that teachers viewed as an ingredient for scaffolding is the character of the interlocutors. As Jane reports in the excerpt that follows, the HP student has to be a well-behaved student who is willing to help his/her interlocutor, and who does not feel superior in any way.

Excerpt 4.5.2 Jane's opinion on the role of character

> *[I prefer] mixed-proficiency pairs, but it depends on who the strong student is; how well-behaved that student is. Because there are students who underestimate [their peers] [...]. Things are not ideal in the school settings anymore [...]. So it depends on the people who work together. And it is better to have mixed-proficiency pairs so that one can help the other [...]. This is the ideal for me.*

Another teacher, Vanessa, referred to a different pre-requisite. She clarified that in her classroom mixed-proficiency pairs only work when the difference in the proficiency level of the two students is not huge. When there is a large gap of proficiency level, the HP student takes over the task and the LP student does not contribute anything. Vanessa suggested therefore that with a smaller proficiency gap, both students would be able to make a contribution. The excerpt below presents her opinion in her words.

Excerpt 4.5.3 Vanessa's opinion on the extent of the proficiency difference

> *There are two things that can happen. If it's mixed-proficiency and the difference is big, then the weaker student will simply wait for the strong student to complete it and will not do anything [...]. There should be some difference so that the group or the pair can cope better with the task, but the level should be close in order for both of them to be able to contribute. Or else, they will sit there and look at each other [...]. This is why I disagree a little bit with mixed-proficiency for all subjects. The mixed-proficiency often forces the weak student to wait for the strong student to finish it [...]. If the difference is big, [the weaker students] will not say anything. Especially if they are shy. Whereas when their level is closer, the result might not be great but at least both of them will make a contribution.*

While recognising that mixed-proficiency groups work well, another teacher, Alison, said that HP students were at a disadvantage, when they worked with LP students. This is because they did not challenge themselves enough and did not push their boundaries. Alison

suggested therefore that when HP students finished working with their LP peers, they should be given additional task targets in order to challenge themselves. This would help avoid losing the interest of stronger students in the task due to the fact that what they do might be too easy for them. The following excerpt presents her opinion.

Excerpt 4.5.4 From Alison's interview

> *For years they said mixed-proficiency, but I find that students who are very strong are at a disadvantage in mixed-proficiency. So I would say mixed-proficiency groups, but for the very strong students there should be an additional target. Some students should leave the groups later and form another group on their own. So that the weak students are helped by the strong students and then [...] the strong students get something extra.*

Many teachers recognised that, while having mixed-proficiency pairs helps in terms of scaffolding, there is also the possibility of HP students taking over the task and completing it on their own without any contribution from their LP peers. In order to avoid this problem, a few teachers said that they prefer mixed-proficiency pairs, but only under the teacher's guidance and observation. More specifically, they added that the teacher should assign specific roles to the students, or ask the HP students to assign roles among them, in order to avoid the lack of contribution on the part of LP students. When LP students own a part that is crucial for the completion of a task, then there is no danger of HP students dominating the task. Alison and Vanessa were two of the teachers who shared this view.

Excerpt 4.5.5 Teachers' opinion on the importance of teachers' supervision

> Alison: *"The strong student can help the weak student, but under the teacher's supervision. When there is no supervision, the strong student will complete it on his/her own and the weaker student will just watch him/her. When there is teacher supervision, the teacher can distribute the activity and the weak student can start for example and the strong student can finish it [...]. The teacher has certain requirements from each student."*
>
> Vanessa: *"[The strong student] can help the weak student when there is guidance from the teacher; when both the weak student and the strong student know what to do. If it's free, the strong student will do all the work and [the task] will be presented from the strong student too; because often they don't even let [the weaker students] present."*

While most of the teachers believed in mixed-proficiency pairs, one of the teachers, Laura, seemed reluctant to provide a definite answer. The reason is because she thought that mixed-proficiency pairs only worked effectively, when the students in a classroom maintained a good relationship between them. She further explained that in order for this to work, HP students should be willing to help their peers without making them feel inferior. The following excerpt shows her response.

Excerpt 4.5.6 Laura's opinion on mixed-proficiency pairs

> *It's a dilemma. I will not give you the easy "recipe" of having mixed-proficiency pairs. Based on my experience, I believe that it depends on what kind of students you are dealing with. Sometimes mixed-proficiency works, but it depends on what kind of classroom it is and what relationship [the students] have developed between them [...]. Sometimes mixed-proficiency pairs may not work because the good student makes the weak student feel inferior. There are cases therefore where mixed-proficiency pairs do not work and [the teacher] should use same-proficiency pairs, or pairs whose proficiency gap is not huge, so that one can gain from the other easily.*

Concluding the mixed-proficiency/same-proficiency pairs section, it is clear that most of the teachers in the present study have a preference for mixed-proficiency pairs because such pairing facilitates scaffolding. Nonetheless, some teachers referred to certain pre-requisites that they thought were necessary for mixed-proficiency pairs to work effectively. These pre-requisites included a well-behaved high proficiency student who would be willing to help his/her lower proficiency interlocutor; a small proficiency gap that would allow both students to contribute to the completion of the task; additional task targets for additional challenge of the HP students; teachers' supervision when organising the task; and finally a good relationship between classmates that would facilitate scaffolding.

4.5.2 Scaffolding

As the majority of the teachers expressed a preference for mixed-proficiency pairs due to scaffolding, I proceeded to ask the teachers about the ways in which they believed scaffolding took place in their classrooms. Teachers referred to a number of ways in which HP students could help LP students. In their response, almost all teachers referred to vocabulary gains. HP students can provide L2 vocabulary when LP students get stuck or they can explain unknown L2 vocabulary to their peers. Other teachers referred to situations where LP students get stuck and HP students provide a keyword that helps the former overcome difficulties. In addition, HP students can help LP students by explaining instructions. Some teachers said that when LP students did not understand exactly what they should do, HP students often explained it to them. These ways in which scaffolding takes place are summarised in the excerpts below coming from Nichole's interview.

Excerpt 4.5.7 Teachers' opinion on ways of scaffolding

> Nichole: *"First of all when the weak student gets stuck and does not know how to respond, the [stronger student] can give him a keyword to help him/her continue and not get disappointed and say "I won't continue because I don't know what to say". Or [the weak student] may not understand exactly what s/he should say and the stronger student can explain: "We have to do this. You have to say this".*

The teachers also referred to the psychological benefits that LP students can have by working with more proficient peers. One of the teachers, Rachel, said that LP students feel safer speaking to their peers, rather than in front of the whole classroom or to their teacher. Other teachers mentioned that LP students feel much safer presenting their part of the task in front of the class, when they have already rehearsed it in front of a more proficient peer. It is as if they have the approval of their more knowledgeable peer that what they will present is correct. So this helps them overcome feelings of embarrassment that occur when they are asked to produce L2 output, while the experience of presenting something without any errors

in front of others boosts their confidence. Jessica is one of the teachers who supports that working with more proficient peers, helps LP students overcome feelings of insecurity and embarrassment. Her words are presented in the excerpt below.

Excerpt 4.5.8 Psychological benefit

> Jessica: *"[The weaker students] feel much more secure when they are in pairs, especially when their partner knows a bit more than they do [...]. There are students who feel uncomfortable speaking or even reading in English because there are students that could make fun of them. They are sixth grade students and they feel this insecurity more intensely."*

Laura also had a similar opinion. She commented that contributing something to the completion of the task made LP students feel that they became a member of the team. So there is a benefit in the sense of socialising with peers. She added that as a result LP students created a more positive self-image. In her words therefore, LP students had more of a sentimental gain when they worked with their HP peers. Another interesting point coming from Laura's response was that LP students accepted and retained more what came from their HP peers, rather than what came from their teachers, because the teacher's comment might be perceived as scolding. In other words, a comment coming from a peer is more positively perceived than one coming from the teacher. Based on this observation, Laura concluded that in general students can have more gains from their peers than their teacher.

The same point was raised by other teachers in their interviews. Catherine for example mentioned at one point that students accept more easily something that comes from their peers rather than their teacher. This is because they feel closer to their peers and they feel much more comfortable with them. Ellie also added that a peer can do a lot more good to LP students than the teacher.

Concluding this section therefore, it seems that HP student can help their LP peers in many ways. They can provide them with the necessary L2 vocabulary and they can help them

understand what they have to do. Most teachers however referred to the psychological benefits of working with more proficient peers. Less proficient students feel safer presenting their part of the task in front of others, when they feel that they have their peers' approval. This helps them overcome feelings of insecurity. Finally, teachers seem to agree that less proficient students seem to gain a lot more from their peers than from their teacher because they are much more receptive when something comes from people they are closer to.

4.5.3 Interlocutor familiarity

Continuing on the subject of pairs, teachers were asked about another issue that emerged in the literature review; namely the effect of interlocutor familiarity on task performance. It has been suggested in the past (e.g. O'Sullivan, 2002) that being familiar to the interlocutor facilitates collaborations and leads to a better result. A number of teachers in the present study mentioned that for practical reasons, i.e. limited time and avoiding confusion during the lesson, students in their classrooms pair with the person sitting next to them. However, despite the practical issue of allowing students to choose their interlocutor, the majority of the teachers did not see a benefit coming from it in any way. They believed that if they offered them this freedom, students would pair with their close friends and it would be more likely for such pairs to end up talking about things other than the task. The excerpts below illustrate this by presenting the opinions of a few teachers.

Excerpt 4.5.9 Teachers' views on interlocutor familiarity

> Melanie: *"[In case they are friends] they might speak about irrelevant topics and produce no result. But when they are assigned with someone else, they are forced to speak about the topic of the task and the result might be better."*

> Rachel: *“I think it is better [for the teacher to assign the pairs] because when they are friends they might start talking about other things.”*

> Victoria:*”Usually they pick their friend, which does not mean that this will be helpful on their performance. Most of them will talk about other things.”*

As shown in the excerpts above, most teachers expressed their belief that pairing with close friends is not a good idea mainly because it becomes more likely for close friends to go off topic and talk about matters irrelevant to the task. In fact, Nichole said that she has tried allowing her students to choose their interlocutors once before, but the result was not good because of the reason cited above; going off topic.

One of the teachers however, Laura, approached this issue from another perspective when she gave her answer. Having initially reported that she supported both options, i.e. teachers assigning interlocutors and students choosing their own, Laura found this to be more of a matter of acceptance. More specifically, she explained that students should not reject their interlocutors simply because they cannot have the perfect dialogue with them. Students should learn to accept their interlocutors despite their level of oral performance. In other words, what Laura suggested is that in real life people do not choose their interlocutors based on their oral skills and this issue of accepting everyone and being able to build a conversation with everyone should be cultivated from a young age. Her opinion is shown in the excerpt below.

Excerpt 4.5.10 Laura’s opinion on acceptance

> *When you speak a foreign language, it doesn’t mean that your interlocutor is always as smart as you, and that you can have a dialogue with him/her the way you want it to be. You will encounter people who might not speak as well as you. So you should learn to accept people who speak the language less fluently than you. So basically it is a matter of acceptance; accepting that there are people who do not have the same skill as you, or have as much fluency as you. You can help your fellow student speak and construct a correct sentence in the sense that the good student takes the role of a guide. But the good student cannot always take this role because it eventually gets boring. This student*

> *sometimes needs to play the role of a student and to be competitive with someone who is equally good. So everything has to be done in moderation.*

It seems therefore that while Laura supports that students should become more tolerant and should be able to help their interlocutors, whoever they are, communicate with them, she also adds that this should be done in moderation. Higher proficiency students, who usually take the role of a guide, need to feel challenged at times by conversing with someone who is of an equal or a higher proficiency level. So the idea of being challenged in order to improve at language learning is also one that emerged from the discussion on proficiency pairing.

Concluding on the issue of interlocutor familiarity, it seems that the teachers do not offer the freedom to their students to choose their interlocutor both for practical reasons (i.e. limited lesson time), as well as pedagogical reasons. The latter involves the teachers' general belief that pairing with a familiar interlocutor does not necessarily have a positive impact on the interaction of the students. In fact, they believe that this freedom will give them the opportunity to pair with close friends, with whom they will most probably discuss off-task topics.

Comparing these views with those expressed by some researchers in the past (e.g. O'Sullivan, 2002; Hancock, 1997), it seems that the two contradict one another, as there is evidence suggesting that learners work better and achieve better results when working with familiar people. In addition, the idea of accepting the interlocutor emerged in the interviews. A teacher suggested that students, as in real life, should become tolerant and accept all kinds of interlocutors they might encounter. However, the idea of being challenged in order to progress in language learning was also proposed. By conversing with interlocutors of an equal or higher proficiency, students are more likely to be challenged and consequently improve.

4.5.4 Conclusions on pair work

This section first explored the teachers' opinions on mixed-proficiency pairs and same-proficiency pairs. It was found that the majority of the teachers prefer for their students to work in mixed-proficiency pairs or groups because this combination facilitates scaffolding. However, some teachers warned that in order for mixed-proficiency pairs to work there are certain pre-requisites that have to be met. These included the character of the students, a smaller proficiency gap, additional task targets for more proficient students, teachers' supervision for assigning roles and the working relationship between classmates.

Furthermore, teachers were asked about the ways in which scaffolding can take place between their students. Their responses referred to vocabulary gains, clarifying by giving instructions, as well as to psychological benefits especially for lower proficiency students. More specifically, psychological benefits of lower proficiency students included overcoming insecurities, socialising and feeling that they become a member of the team by contributing something in the completion of the task. Overall, it emerged through the teachers' responses that students can gain a lot more from their more proficient peers than from their teachers.

The final section of the discussion on pairs concerned interlocutor familiarity. From the teachers' responses, it emerged that teachers do not allow their students to choose their interlocutor for practical reasons, as well as pedagogical reasons. More specifically, they believed that this freedom can often result in off-task talk, as the students would choose someone they feel closer to. An interesting point that emerged from the data was that students should be able to accept and communicate well with any one of their peers because, as in real life, people do not get to choose their interlocutors.

4.6. Chapter summary

This chapter presented the qualitative analysis of 19 teacher interviews, in an attempt to provide a response to the first research question on teachers' views on four main areas of the task-based learning approach. These topics, which also formed the structure of the chapter, concerned task implementation in Cyprus, task repetition, the potential of using metalanguage in the L2 and the role of pair combinations in terms of student proficiency.

Regarding task implementation first, teachers seemed to view tasks as a very valuable addition when teaching a foreign language because of their pedagogical, practical and psychological benefits. Due to this variety of benefits, communicative activities which involve pair and group work, and not tasks in the sense that Samuda and Bygate (2005) intended, were very often employed in this context. These communicative activities, which were included in the students' book, did not allow for much creativity to take place because students of all proficiency levels should be able to participate in all activities going on in the classroom. This brought to the surface the issue of mixed-proficiency classrooms, which was caused by the fact that the majority of the students attended private lessons in English from a very young age.

It seemed therefore that the extent of creativity of the tasks that each classroom completed depended on the teacher. From this discussion, other issues that concerned teachers emerged. These included their lack of time to adequately prepare for all the subjects they teach, their lack of expertise, and consequently confidence, to teach foreign languages and their need for new innovative ideas for creating their own materials. In general, it emerged from the interviews that there were a number of factors that should be considered in order to answer the question of task implementation in this context.

In terms of task repetition, the teachers in general expressed their preference for repeating the same types of tasks, but not with the same content, because students often lose interest in activities that they have already completed. The repetition of the same types of tasks gave the opportunity to students to focus more on the language, as they were familiar with the structure and the aims of the task. Nevertheless, teachers mentioned repeating exactly the same tasks only to refresh their students' memories from the previous lesson before the introduction to the new lesson. However, they pointed out that same-task repetition only had psychological benefits for lower proficiency students, whose confidence was boosted when they contributed to task completion. No benefits were reported for higher proficiency students.

The third matter of discussion in the interviews concerned metalanguage. All teachers agreed that metalanguage in this context always occured in the L1. While teachers in the Cantonese context (Carless & Gordon, 1997) felt that this use of the L1 reflected their own 'failure', teachers in the context of the present study believed that this was the result of limited time and mixed-proficiency. Nevertheless, many placed their hopes on the new curriculum, which was expected to contribute a lot more to the development of the oral production skills of learners. Under those circumstances, teachers expected that future fifth and sixth grade students would be in a better position to use metalanguage in the L2.

Finally, the teachers were asked about their opinion on the effectiveness of different combinations of pairs. The analysis of the data showed that in general teachers preferred mixed-proficiency pairs because these allowed for scaffolding to take place between students. Teachers reported that scaffolding came in the form of vocabulary gains, clarifying instructions as well as psychological benefits for the less proficient students. However, in mixed-proficiency pairs, there was always the danger of HP students dominating the task and LP students contributing nothing. Some teachers therefore referred to some pre-requisites for

effective mixed-proficiency collaborations. These included the character of the students. Higher proficiency students should be willing to help their peers; also, a relatively smaller proficiency gap, as well as additional task targets for more proficient students; the teachers' supervision for role assignment; and a generally good relationship between classmates.

The section on pair combination ended with reference to the role of interlocutor familiarity. Choosing their interlocutor was not often the case in this context for practical, as well as for pedagogical reasons. The value of interlocutor familiarity was not acknowledged by these teachers, so it could be argued that in general this result contradicted with past research evidence suggesting otherwise (e.g. O'Sullivan, 2002). An interesting point that was raised, however, was that students should not always be allowed to choose their interlocutors because they should learn to accept all kinds of interlocutors, as it happens in real life.

The chapter served as a more in-depth exploration of the context of the study. The results presented here are discussed in more depth in relation to those of previous research in Chapter 7.

CHAPTER 5: Qualitative analysis of student interactions

5.1. Introduction

The present chapter attempts to provide an answer to the second research question, which addresses the following inquiry: *How does the output change from session 1 (S1) to session 2 (S2) particularly in terms of content?* This inquiry was mainly based on Bygate's (2001) argument favouring task repetition. As stated in the literature review, Bygate (2001) maintained that when repeating a task speakers have the opportunity to focus their attention on different aspects of speech, such as fluency, accuracy and complexity, as they have already formulated the necessary structures on the first occasion. However, building on his argument, the present study investigates TR+; so what is investigated in this chapter is the change of content after repeating a task with feedback.

As the effects of TR+ on two of the components of oral production, namely fluency and accuracy, are already investigated in the present study through the quantitative analysis presented in the following chapter, this chapter focuses on the content of students' output. Content was chosen here over complexity, which is the third component of oral production, because based on the hypothesis of the present study content is the one that is likely to change. More specifically, with increased WTC, students are expected to attempt to produce more complicated ideas when they repeat the task. This chapter examines qualitatively whether and how the spoken output changes from the first to the second occasion of completing the task; in other words, whether the output is 'enriched' in any way. The concept of richer output here includes instances where additional ideas are expressed; instances where existing ideas are extended or become more complicated; and instances where a wider range of vocabulary is used on the second occasion.

5.2. Data and sample

The data for this analysis come from the student interactions taking place during the 8-week intervention. Over the two sessions of each week, the students of the two experimental groups completed the same task twice and they were being recorded. As this question focuses on TR+, the interactions of the students of the comparison group were not considered in this analysis.

The interactions of pairs who remained together on both sessions of each week were transcribed and considered for the analysis. As discussed in the literature review chapter, the reason pairs who changed interlocutor on the second occasion were excluded is because the interlocutor familiarity variable is considered to be influential on the progress of the conversation (O'Sullivan, 2002).

The transcripts were divided based on two criteria. First, they were divided by conditions, namely the CS condition and the EO condition. Second, they were divided by the proficiency level of each pair. As discussed in the Methodology chapter, the students were allocated in a *high proficiency group (HP)* or a *low proficiency group (LP)* based on their C-test scores. As they chose their own interlocutor, three kinds of combinations were created: *HH pairs,* which consisted of two HP students; *HL pairs,* which consisted of a HP student and a LP student; and *LL pairs,* which consisted of two LP students. The most common combination was that of a HL proficiency pair in both groups. The analysis in this chapter also takes into consideration the proficiency match of each pair.

5.3. Qualitative analysis results

Having reviewed all transcripts, enriched output was considered both in the form of *task language,* or the language used as part of the task, and *L2 metalanguage,* or the off-the-record language used for discussing the task. Any kind of evolution of task language or L2 metalanguage from session 1 (S1) to session 2 (S2) was considered as evidence of enriched output. For both types of language, examples from the transcripts of both the CS and the EO groups are examined.

In this chapter, although the focus of the discussion is on task language and L2 metalanguage, reference is also made to three other types of language appearing in the transcripts. The first concerns *L1 metalanguage* and the functions it serves during task completion. The second concerns one of the functions often served by metalanguage, namely *scaffolding*, or in other words, the help that students provide to their partner when completing tasks (Brooks & Donato, 1994). The third concerns *feedback on accuracy,* which is analysed based on transcripts coming from the EO group only and attempts to determine whether this kind of feedback is used by the students. Although these three types of language do not directly contribute to answering the second research question, it was decided to include them in the discussion because they are issues considered in the literature of learner codeswitching or issues that emerge in the transcripts.

The rest of this section focuses first on task language and then on L2 metalanguage, and discusses how these evolve from S1 to S2. The arguments made are illustrated with transcripts as examples. After each transcript, a short discussion on L1 metalanguage, scaffolding and accuracy feedback, where appropriate, follows. All names used in the transcripts are pseudonyms, which are used in order to protect the identity of the participants.

5.3.1. Task language

This section examines whether the task language of both the CS and the EO group was enriched from S1 to S2. Having gone through all transcripts, it seems that there are examples of task language both getting richer and poorer.

i. *Examples of richer task language on S2*

Looking at examples where task language became more creative, the following transcript shows the interactions between a HL pair from the CS group. This pair consisted of Gary, who is an HP student, and Peter, who is an LP student. Their interactions took place during the two sessions of week 3, when they had to complete a narrative task. They were presented with a sequence of five pictures and they were asked to narrate the story in the past. The underlined utterances on S2 show how their output evolved from S1.

Transcript 5.3.1 Example of enriched output from CS group on week 3: Gary (HP) & Peter (LP)

Session 1

1 G: Yesterday Tom, Sally and Lilly saw a poster with a circus on it.
2 P: Then, they went to a bus station, get a bus, I get a bus and go to circus.
3 G: “Circus” [correcting pronunciation]
4 P: Circus.
5 G: Then, they got out of the circus and went to a café.
6 G: “They had an ice-cream”.
7 P: They had an ice-cream.
8 G: “Ice-cream”. [correcting pronunciation]
9 G: And then the clowns get out of the circus, on a table and started dancing.

Session 2

1 G: Yesterday Tom, Sam and Lilly saw a poster with the circus on it. There were some lions, elephants. There were very interested about it.
2 P: They liked it very much and they and they want to...
3 G: “Went to the circus”.
4 P: And they and they wa...
5 G: “Went”
6 P: Went to a circus.
7 G: “To the circus by bus”.
8 P: To the circus by bus.
9 G: The circus wasn’t very expensive. Ten pounds for adults and five for children. But it was very, very nice.
10 P: Then after they want to...they went to the coffee...
11 G: “To a coffee”; Στον καφέ; *[At the coffee]*
12 P: To a coffee...
13 G: “To a cafe”.
14 P: To a cafe and get an ice-cream.

15 G: "Had an ice-cream". Then the clowns from the circus came out and started dancing on the...on the table. Can you believe that? Two clowns on the table.

Before discussing the evolution of task language, it is worth discussing the issue of *scaffolding* in the above transcripts. As the two students differ in terms of proficiency, it is not surprising that Gary helped Peter at several instances. On S1, Gary tried to help Peter by correcting his pronunciation of L2 words on lines 3 and 8. The most common type of help that he provided, however, was by giving him, or rather whispering to him, the exact phrases that he should use next. This kind of help occurred on line 6 of S1, and then lines 3, 5, 7 and 13 of S2. In addition, when Peter used the erroneous utterance "they went to the coffee" on line 10, Gary gave him a chance to reconsider what was wrong on line 11. He repeated the erroneous utterance and then translated its exact meaning to Greek, giving Peter the chance to spot his error. When Peter failed to understand what was wrong with his phrase (line 12), Gary gave him the correct phrase (line 13).

In terms of task language, enriched output occurred in the form of new additional ideas, which came mainly from Gary, the HP student. On line 1, he imagined how the circus would be by referring to lions and elephants. On line 9, he described one of the pictures in much more detail than in the first occasion by referring to the price of the tickets, while on line 15 he tried to convey the surprise of the people in the story ("Can you believe that?"). In addition, Peter, the LP student, tried to be more creative by commenting on the enthusiasm of the people in the story on line 2. In the above example therefore, it is clear that task language becomes more creative when the task is repeated.

Another example of TR+ appearing to have a positive impact on task language comes from the EO group between a HL proficiency pair: Patrick, who was the HP student, and Tom, who was the LP student. On week 1, the students were asked to complete a narrative task. They had to describe to their interlocutor what they did on the previous day, by following a

sequence of pictures. The underlined utterances demonstrate enriched output, which comes in the form of new additional ideas, as well as existing ideas that were more elaborated.

Transcript 5.3.2 Example of enriched output from EO group on week 1: Patrick (HP) & Tom (LP)

Session 1

1 P: Ξεκινώ. *[I start]* Yesterday morning I went to school and had breakfast there. Yesterday eee...yest....yesterday....
2 T: "In the lunchtime"
3 P: Yesterday in the lunchtime I ate lunch. I ate spaghetti with tomato sauce. After lunch I do...I did my homework and sleeping....and sleep...eee...half...half hour. In the afternoon I went to the park with my friend Tom and we played football. In the evening I ate dinner and I...and I watch TV. After lunch I did a bath....I did a bath and...I did a bath. At eleven...eleven...eleven o'clock I was sleeping.
4 T: Yesterday morning I have breakfast in my home. Yesterday lunchtime I...I ate....I ate pizza. Yesterday afternoon I went to the park with my friend Patrick and played football. Yesterday evening I have my bath and go to sleep at 9 o'clock.

Session 2

1 P: Yesterday I woke up in the morning at...at ten o'clock and I went to brush my teeth. I ate a breakf....I ate breakfast...I ate breakfast. I ate a sandwich. Then I went out to help my mum with the garden. I had lunch and at two o'clock...two o'clock I ate spaghetti. In the afternoon I went to the park with my friend Tom and played football and basketball. I...I came back home at...at...*[pause for 2 seconds]*
2 T: "Seven o'clock"
3 P: Seven o'clock and I went to a restaurant with my family. I ate ni...I ate pizza.
4 T: "For dinner"
5 P: For dinner. I went to bed at eleven o'clock. That was my day.
6 T: Yesterday I wak...I woke up [reform.] at nine o'clock. After I go with my family....I went with my family to a restaurant to eat for lunch. I ate hamburger and drink coca-cola. In the afternoon I went with...with my friends to the park....with my friend Patrick and Henry and play football and tennis. I...e...in the evening I...I sleep at ten o'clock. This is my day.
7 P: "That was my day".
8 T: That was my day.

There are instances of scaffolding in Patrick's and Tom's conversations. These are mainly expressed in a more direct form; that is they provide each other with the exact phrase that their partner should use, when he is stuck. Such instances occurred on line 2 of S1 and lines 2 and 4 of S2. What is interesting here is that it is the LP student, namely Tom, who provided these phrases to his interlocutor. This proves that in this case, the HP student did not take

over the task and the LP student was able to contribute to completing the task. The final instance of scaffolding came from Patrick on line 7 of S2, when he corrected Tom's erroneous utterance by providing the correct version.

Before looking at task language, it is worth referring to Patrick's opening line on S1, which was in Greek. As the roles were already set for them, Patrick used a Greek word (equivalent to "I start") to signal perhaps the beginning of the task. This was the only instance of codeswitching during their interactions.

As this pair was under EO conditions and received feedback on accuracy before completing the task for the second time, reference to any impact that feedback might have had should be made here. As the handout with the feedback contained the correct versions of errors made by the whole group, only two of the errors made by this pair were corrected in the handout. These were Patrick's utterance "I did a bath" on line 3 and Tom's utterance "I have breakfast" on line 4. While the idea of having a bath was not used on the second occasion, the phrase "I ate breakfast" which is still correct was used, instead of "I had breakfast" on S2. It cannot be argued therefore that this case demonstrated intake of accuracy feedback.

In terms of task language, it seems that this pair enriched their output at S2 by adding new ideas. Although they were given pictures and some prompt words for guidance, both groups were instructed to be creative if they wished to. Having skipped a few pictures, this pair tried to be innovative during their interactions. While Patrick, the HP student, was creative at S1 as well as S2, Tom, the LP student, showed an increase of creativity at S2. Some of the new ideas he expressed were already used by his interlocutor. For example, on line 6 of S2 Tom added that he went to a restaurant with his family; an idea previously expressed by Patrick on line 3 of the same session. As this was not a prompt provided by the task, it seems reasonable to suggest that Tom benefitted from the creative output of Patrick. This is illustrated by the

fact that he did not just borrow Patrick's idea, but he built on it, by commenting on the food and drink he had (line 6). It seems therefore that LP students can sometimes benefit from their HP peers in terms of creativity.

Another example showing increased creativity in terms of task language from S1 to S2 is included in Appendix 34.

ii. Examples of poorer task language on S2

Having discussed examples of pairs who produced richer content on S2, there are also examples of pairs who produced poorer output on S2 than on S1. The following example comes from the interactions between Gary and Peter from the CS group on week 6. On that week, they were asked to complete a doctor-patient roleplay, by following the pictures. Transcript 5.3.3 below presents the conversations that took place. As in the other transcripts, the underlined phrases show utterances that were not prompted by the task itself. In other words, they show some kind of creativity.

Transcript 5.3.3 Example of reduced creativity at S2 from CS group on week 6: Gary (HP) & Peter (LP)

Session 1

1 P: Εγώ είμαι ο γιατρός. Άτε εσύ ρώτα. Εσύ είσαι ο ασθενής. Εγώ είμαι ο γιατρός. *[I am the doctor. You ask. You are the patient. I am the doctor.]*
2 G: [he coughs] I'm very ill today. I will go to the doctor. Hi doc.
3 P: Hi.
4 G: "Hi Gary".
5 P: Hi Gary. How are you doing?
6 G: I'm very sick doctor. "That's very bad".
7 P: Oh, that's very bad.
8 G: Yes, I know it. Doctor should I go to the school tomorrow?
9 P: No, you shouldn't go to the school. You...you...
10 G: "You must stay home"
11 P: You must stay home.
12 G: "Lying in bed"
13 P: Lying in bed.
14 G: Ok. Should I go for a walk with the dog?

15 P: No, you shouldn't go walk with the dog. You should drink... *[pause for 2 seconds]*
16 G: "Eat a soup"
17 P: It's soup. A hot soup.
18 G: Soup. Yak.
19 P: Πως ένι το εννα κάμω εμετό; *[How do we say "I'll throw up?"]*
20 G: I will throw up. Should I play football with my friends?
21 P: No, you shouldn't play football with your friend. You should drink a medicine.
22 G: Yak. That's worse than soup.
[they whisper]
23 P: "You have to see if you have a temperature".
24 G: Must I take my temperature doctor?
25 P: What?
26 G: "Yes." Πε yes. *[Say yes]*
27 P: Yes.
28 G: Should I go swimming?
29 P: No, you shouldn't go swimming. You should wear a jacket.
30 G: Οι ρε "jacket". Πως το λέμε; *[Not "jacket". What's it called?]*
31 P: Σακάκι. *[Jacket]*
32 G: Ok. But why? I don't want to lose...to lose swimming lessons [he coughs]
[they whisper]
33 P: If you want to be more sick, then go.
34 G: Ok. Bye.

Session 2
1 G: Hi Peter.
2 P: Hi doctor.
3 G: You...are you ill?
4 P: Yes, very much.
5 G: Well..."Should I".
6 P: Should I go to school today?
7 G: No, you mustn't. "But we are playing today".
8 P: But we are playing football today.
9 G: I said not. You must stay in your bed....you must stay lying on your bed.
10 P: Should I go...I go my dog for walk?
11 G: No, you must not. You shouldn't get...go out of home. You should...you should eat all your soup.
12 P: Should I go...should I play football?
13 G: Of course not. You...you must drink your medicine and you must count your temperature.
14 P: Should I go for swimming?
15 G: No, you must wear your coat until you...until you feel better.

Similar to their conversations on week 3, Gary provided a lot of support to Peter on week 6 as well. On several instances (lines 4, 6, 10, 12, 16, 23, 26 of S1 & lines 5, 7 of S2) Gary

provided Peter with the exact phrases that he should use. This shows that Gary was taking control of the task and guided Patrick through the conversation.

Furthermore, it is worth looking at the role of their L1 use on S1, which was mainly used for metalanguage. Their first switch to Greek came at the beginning of S1, when Peter chose the role he wished to play in the task and allocated the remaining role to Gary. Then on line 19, Peter switched to the L1 again in order to ask Gary about the English equivalent of a certain phrase (How do we say "throw up"?). The next instance of L1 use occurred on line 26, where Gary instructed Peter to reply positively to his question. This was probably because Gary realised that Peter either did not understand what Gary had asked or he did not hear him clearly. In order to speed things up perhaps, Gary chose to direct his interlocutor. Finally, the pair switched to Greek on lines 30 and 31, in order to talk about the L2 equivalent of the Greek word of "jacket". In terms of S2, no L1 switches occurred. To sum up therefore, the pair used L1 metalanguage in S1 for different purposes.

Looking at task language now, as shown from the transcripts above, S1 was longer than S2 and it had a lot more creative utterances. In S1, lines 2 to 6 presented an introduction to the dialogue between the patient and the doctor, which made their conversation more authentic. Then on line 17, Peter added the adjective "hot" to the noun "soup" that had been provided to him by Gary. Later, on line 19, Peter contributed to the creativity of the conversation by prompting Gary to comment that the soup would make him "throw up" (line 20). On line 32, Gary is creative by pretending he is frustrated about having to miss his swimming lessons, which prompted an equally creative response from Peter on line 33.

These creative ideas seem to be reduced on S2 however. While the authentic introduction is maintained, the following lines do not show many creative utterances. A possible explanation could be that students often lose interest when they are asked to repeat the same task.

Although these students were given a relatively valid explanation of why they have to repeat the tasks (i.e. in order to see whether they can use the feedback on S2), their frustration was still evident. Having kept a personal logbook during the period of the intervention, this frustration was noted several times on the part of the experimental groups. This was also confirmed in the teachers' interviews discussed in the previous chapter. Most of the teachers reported that young learners never enjoy repeating the same activities. It should not be surprising therefore that their reduced enthusiasm sometimes led to a reduced effort in completing tasks again.

Another example of reduced creativity is reflected in the interactions of a pair from the EO group. The pair consisted of Gina, who was an HP student, and Mandy, who was an LP student. The two girls completed an information-gap task, which was given to them on week 2. The underlined phrases on the transcripts below highlight utterances or ideas that were not prompted in the handout of the task, and therefore indicate creativity.

Transcript 5.3.4 Example of reduced creativity at S2 from EO group on week 2: Gina (HP) & Mandy (LP)

Session 1

1 G: Let's start.
2 M: Did you go swimming yesterday?
3 G: No, I didn't. I was in the computer lessons.
4 M: Did you go to play tennis yesterday?
5 G: No, my mum took me to the doctor.
6 M: Did you go to play bowling?
7 G: Yeah, that's what I did. At the mall, of course.
8 M: That's great. That's great.
9 G: Now I ask the questions. Did you make biscuits yesterday?
10 M: No, I made yummy cake.
11 G: "A yummy cake"
12 M: "A yummy cake". Ok.
13 G: Το λοιπον. *[Well]* What did you eat for dinner?
14 M: I ate spaghetti.
15 G: I ate pizza. Did you do your homework?
16 M: Of course.
17 G: In the evening.
18 M: Of course.
19 G: Me too.
20 M: That's great.

Session 2

1 M: Let's start.
2 G: Yes. Did you go swimming yesterday?
3 M: No. I went to a computer lesson.
4 G: Did you go...play tennis?
5 M: No. I went to the doctor.
6 G: Did you play bowling?
7 M: Yes. I played bowling.
8 G: <u>It was fun right?</u>
9 M: <u>Yeah.</u>
10 G: Your turn.
11 M: Ok. My turn. Did you make biscuits yesterday?
12 G: No. I made a cake.
13 M: Did you eat pizza yesterday?
14 G: No, I ate spaghetti.
15 M: Mmm yummy. Did you do your homework yesterday?
16 G: Yes, I finished my homework.
17 M: <u>Excellent.</u>
18 G: Thank you.

The only instance of scaffolded help occurred on line 11 of S1, where Gina repeated the correct version of Mandy's utterance, by adding the article "a". As for L1 use, the pair did not switch codes very often, as they were under the L2-only condition. Gina switched to Greek only once on line 13 of S1 with a short phrase, which is equivalent to the English "well" when this is used as an interjection. As this short switch did not contribute directly to the progress of the task, it is considered to be a type of externalised *inner speech* (Antón & DiCamilla, 1999). Perhaps at that moment, Gina was trying to re-direct herself on where she was on that task; trying to find which question she should ask next. Finally, in terms of accuracy feedback, the only error made in S1 was made by Gina on line 3, who said "I was in the computer lesson" instead of "I was at the computer lesson". This was not included in the accuracy feedback, as it was a minor mistake made only by a single student.

In terms of task language, the transcripts above show that the girls had a more creative conversation at S1. Gina, who was the HP student, initiated creative utterances at several points in the conversation. On line 5, she added that she was taken to the doctor by her mother. On line 7, she elaborated her answer by adding that she played bowling at the mall. On line 17, she clarified the time of doing her homework. When the pair switched roles on S2

however, Gina initiated creative utterances only once on line 8, when she asked whether bowling was fun.

Appendix 34 includes a third example of an interaction which has its creativity of task language reduced on S2.

As in the first example, the reduction of creativity at S2 could be due to the loss of interest when repeating the same tasks. Nevertheless, another explanation could be that these two students feel that they know the language that should be used in this task very well, in other words they are in control of the task, and therefore do not experiment with it any more. The idea of being challenged when completing a task in order to be effective was expressed by one teacher in the interviews and could be linked to this example. When learners feel that they are working on something that they already know well and that they are not challenged, then they might not put their full effort into completing the task as well as they can. However, one could argue exactly the opposite, suggesting that learners are more likely to experiment with language that they know well because they feel a lot more comfortable with it. In this case however, it seems that the most plausible explanation would be that, because they lost interest, they did not experiment with the language further.

5.3.2. L2 Metalanguage

This section examines whether TR+ led to enriched output in the form of L2 metalanguage. In order to investigate whether L2 metalanguage evolved in some way, this section is divided in terms of the two experimental conditions, as these are likely to have played a role in the way L2 metalanguage was used in each group. First, this section focuses on the use of L2 metalanguage by the CS group, considering the fact that this group was allowed to switch to

the L1 at S1, but was later introduced to and encouraged to use L2 metalanguage at S2. Then, the use of L2 metalanguage by the EO group is examined, in relation to the L2-only rule. As in the previous section, after the transcripts reference is made to instances of L1 metalanguage, instances of scaffolded help and the implementation of accuracy feedback, where appropriate.

i. L2 metalanguage of CS group

Looking at the CS group first, five types of L2 metalanguage instances emerged in the transcripts. These are the following:

a) L2 metalanguage phrases at S2, which appear on the handout of that week.
b) L2 metalanguage phrases at S2, which appear on the handouts of previous weeks.
c) Creative L2 metalanguage at S2, which does not appear on any of the handouts.
d) L2 metalanguage phrases at S1, which appear on the handouts of previous weeks.
e) Creative L2 metalanguage at S1, which does not appear on previous handouts.

One of the interactions that incorporated instances of all five types of L2 metalanguage was the interactions between Chris and Ivy, who were a HH proficiency pair. Their conversations took place on week 4, when the pair was asked to complete a narrative task following the story shown in a sequence of five pictures. The following transcripts present their conversations. The underlined utterances are all the instances of L2 metalanguage.

Transcript 5.3.5 Example of types of L2 metalanguage of CS group on week 4: Chris (HP) & Ivy (HP)

Session 1

1 C: Ήνταλως εν η γόμα? *[How do we say glue?]*
2 I: Ποια γόμα? *[What glue?]*
3 C: Glue?
4 I: Εγω να πω τούτη, τούτη και τούτη? *[I will talk about this one, this one and this one.]* Αρκεφκεις. *[You start]*
5 C: What did Billy do yesterday? Εγώ αρκέφκω. *[I start]*
6 I: Ντα. Κανονικά όμως. *[Ok. We start properly.]*

7 C: Νταξει. *[Ok]* We start now really.
8 I: Αρκεφκουμε τωρά αλήθεια. *[We start now really.]*
9 C: Ναι, ναι ναι. *[Yes, yes, yes.]* Yesterday Billy was playing with his ball.
10 I: Τι εννα πω δαμε? *[What should I say here?]*
11 C: "But..."
12 I: But...but...but the ball fell on the table and broke...and broke the table.....and broke it.
13 C: Billy tried to glue the...the...the table.
14 I: The cake?
15 C: Ναι. *[Yes]*
16 I: And then his mother made...made him a cake.
17 C: But unfortable the table broke...
18 I: "When he put the cake on the table".
19 C: Sorry, sorry, sorry, sorry, sorry. But when...
20 I: "His mother"
21 C: His mother put the cake on the table, the table broke again.
22 I: "Break".
23 C: Broke again and...and the story finish.
24 I: Και *[And]* "The mother's angry".

Session 2

1 C: Now we have to do it in English. One picture for you and one for me. So three for you and two for me. You understand? You understood?
2 I: Yes I understood.
3 C: You you say that that that and I say that that. *[pointing at pictures]*
4 I: Οι. *[No]* You'll say this...
5 C: No, no. That that.
6 I: Yes.
7 C: No.
8 I: Yes. Οι εγώ τούτα τα τρία. Οι. Εσύ τούτα τα τρία. Τέλειωνε. *[No, I'll talk about these three. No. You talk about these three. Go on.]*
9 C: I start.
10 I: I start.
11 C: I start.
12 I: Ντάξει. *[Ok.]*
13 C: What did Billy do yesterday? Yesterday Billy played with his ball.
14 I: "In the saloon".
15 C: In the...
16 I: "Saloon"
17 C: Έννεν saloon. *[It's not "saloon"]*
18 I: Εν saloon. Τι ένι; *[It's saloon. What is it?]*
[They ask me and I say it's the "living room"]
19 C: Billy was playing with his ball in the living room.
20 I: After that the ball fall fell on the table and break it.
21 C: The Billy ee
22 I: Ρώτα το τούτο; Ρώτα το τούτο; *[Ask this one. Ask this one.]*
23 C: How do we say in the past tense?
24 I: "How do we say glue in the past tense?"
25 C: Τέλοσπάντων. *[Anyway]*
[pause]
26 C: Εκάμαμε τα σαλάτα. "Προσπάθησε" ήνταλως ένι; *[We messed it up. How do we say "try"?]*
27 I: "Try".
28 C: Billy was...now we start. Really start...Billy was playing with his ball in his living room.
29 I: After that the ball fell on the table and break it.
30 C: Billy tried to glue the table.
31 I: And he did it. After that his mother came to the living room...
32 C: "With a cake"
33 I: With a big big cake.

34 C: When when his mo...when his mother....
35 I: “Put”
36 C: Put the cake on the table, the table broke again.
37 I: And then his mother was very very angry.
38 C: Ok. Now we finish.

In the above transcripts, there are a few instances revealing scaffolded help between Chris and Ivy. On lines 11, 18 and 20 of S1, and on lines 14, 16, 24, 32 and 35 of S2, Ivy, on most of these instances, gave Chris direct help, by giving him the phrases that she thought he should use next. Also, on line 22 of S1, Ivy thought that Chris’ use of the verb “broke” was wrong. She attempted to correct him therefore by giving him what she thought was the correct version (i.e. “break”). Chris did not follow her suggestion, as on line 23 he correctly insisted on the verb “broke”.

Before examining the evolution of L2 metalanguage, it is worth looking at the functions that the L1 metalanguage served in the above transcripts. Looking at S1 first, Chris and Ivy started their conversation with Chris asking in the L1 about the L2 equivalent of the word “glue”. Using the L1 in order to find L2 equivalents of L1 words is one of the functions that have already emerged in the discussion of previous transcripts in this chapter. Then, Ivy continued using Greek in order to organise the task and, more specifically, allocate the pictures that each one of them would talk about (line 4). At the end of that line, she informed Chris that he should start narrating. Having read the guiding question of the task (“What did Billy do yesterday?”), Chris confirmed in Greek that he would start the narration (line 5). When Chris finished his description of the first picture, Ivy switched to Greek in order to ask for Chris’ help on what she should say about the second picture (line 10). Chris provided help by giving her the first word that she could start with (i.e. “But”), hoping that this would prompt her response.

Instances of L1 metalanguage appeared on S2 as well. Having had an argument about the allocation of pictures, Ivy stressed which pictures she wanted to talk about by switching to Greek on line 8. Then on line 17 and 18, Chris switched to the L1, in order to show his disagreement about an L2 word that Ivy had used. Ivy asked in Greek what the L2 equivalent of the word "living room" was. Later on in the conversation, Ivy switched to the L1 in order to show her partner what he should ask next (line 22). Chris followed her recommendation. Finally, on line 26, Chris' L1 use served two functions. At the beginning of the line, Chris commented on their performance thus far by saying that they messed it up. Then he remained in Greek and asked Ivy about the L2 equivalent of the word "try". This discussion of L1 metalanguage therefore demonstrates that the L1 was used for several purposes in an attempt to complete the task.

Concerning L2 metalanguage, the transcripts cited above revealed that L2 metalanguage was used on several instances during Chris' and Ivy's conversations. In regards to S1 first, although they were allowed to switch to the L1 whenever they needed to, there were a few instances where L2 metalanguage was used, sometimes creatively and others as a variation of L2 metalanguage phrases previously given. For example, Chris says on line 7 "We start now really". This is a variation of phrases that appeared on the handouts of the three previous weeks. These handouts included the phrases "You can start" and "You start", on which Chris could have based his comment. Moreover, there are examples of L2 metalanguage being used creatively in this session. On line 19 Chris apologised in English for omitting a part of the story, while on line 23, he marks the end of the story by commenting in the L2 that the story has now finished.

In terms of S2, as it was expected, more instances of L2 metalanguage occurred. Some of them came from that week's handout. For instance, on line 1 Chris used the phrases "Now we have to do it in English" and "One picture for you and one picture for me", which came from

that week's handout. In addition, on lines 9 to 11, they were trying to decide who is going to begin by using the phrase "I start", which again was included in that week's handout. On line 23, Chris again used a phrase appearing on the handout in order to ask for the L2 equivalent of a Greek word. Finally, worried that their performance was not good enough, Chris reported on line 28 that this was the moment I should consider as the beginning of their conversation. For this, he used a variation of previous L2 metalanguage "Now we start. Really start."

What is interesting is the creative L2 metalanguage that is being used in S2 by this pair. On line 1, having used a couple of phrases from the handout, Chris became more creative. He tried to ensure that Ivy understood what they have to do, by further explaining that she had to talk about three pictures and that he would talk about two pictures. Although not error-free, he then added: "You understood?". His question prompted Ivy's response in the L2 on line 2 "Yes, I understood". Then on line 3, Chris continued to speak in the L2 when he allocated three pictures to Ivy and two to himself. On lines 4 to 6, the short argument that followed between them, regarding their allocated pictures, occurred in the L2. Finally, on line 38, Chris marked the end of their conversation with the comment "Now we finish". This did not appear on the handout of that week or of previous weeks, so it is considered here as part of Chris' creative metalanguage.

Closing the discussion on this example, it seems that, while Chris and Ivy experimented with the feedback on L2 metalanguage appearing on the handout of either that week or of previous weeks, they also made an effort to remain in the L2 for metalanguage by being creative. This could suggest that the idea of using metalanguage in the L2 in this context is not as far-fetched as the teachers in this context perceived it to be in their interviews. Another interpretation could be linked to the conclusions of DiCamilla and Antón (2012), suggesting that these learners begin to regard the L2 as a system that can be used for learning, rather than

a system to be learned. Nevertheless, the important observation here is that there are students who are able to at least make an effort to remain in the L2 for metalanguage. In this example, learners did not just use the examples from the handout. They experimented with the feedback and they were being creative with this kind of language on both sessions of subsequent weeks, showing thus real control of the language. This supports the part of the hypothesis suggesting that the recycling of L2 metalanguage will lead to increased L2 practice.

Another good example of the different types of L2 metalanguage used by the CS group is the conversation that took place between Amy and Laura on week 6. This was a HL proficiency pair, with Amy being the HP student and Laura the LP student. On week 6, they were asked to complete the doctor-patient roleplay task. The following transcript presents their conversations. Again the underlined utterances represent L2 metalanguage.

Transcript 5.3.6 Example of types of L2 metalanguage of CS group on week 6: Amy (HP) & Laura (LP)

Session 1

1 A: Είσαι ο *[Are you]* student A or are you student B?
2 L: I am student...student A.
3 A: Ok. E...Ξεκίνα. *[Start.]*
4 L: Should I go to school?
5 A: No, you shouldn't go to school. You should stay in bed.
6 L: Should I walk the dog?
7 A: No, you shouldn't walk the dog. You should drink some...eat some soup.
8 L: Should I...I play football?
9 A: No, you shouldn't play football. You should have some medicine.
10 L: Πως το λένε; *[What's it called?]* [Shows the picture with the thermometer to Amy] Should I...
11 A: "Get my temperature"
12 L: *[No answer]*
13 A: "Get my temperature"
14 L: Get my temperature?
15 A: Yes, you should.
16 L: Should I go to swim?
17 A: No, you shouldn't go swimming. You should wear a jacket.
18 L: Ετελειώσαμε. *[We're done.]*
19 A: Ετελειώσαμε. *[We're done.]*

Session 2

1 L: Πε εσύ. *[You say.]*
2 A: Should I go to school? A. Student A is the patient. Student B is the doctor. I want to be student A. You are the doctor.
3 L: Ωραία. *[Good.]* Let's start.
4 A: Should I go to school?
5 L: No, you shouldn't go to school. You should stay to bed.
6 A: Should I walk the dog?
7 L: No, you shouldn't walk your dog. You should eat some soup.
8 A: Should I play football?
9 L: No, you shouldn't. What does this picture show?
10 A: This shows a medicine bottle.
11 L: [She taps on the handout. She wants to show Amy something.]
12 A: Εσύ πρέπει να πεις. *[You have to say something.]*
13 L: Ενιξερω πως το λεμε. *[I don't know how we say this.]*
14 A: You should ask me if I...you should tell me to drink some medicine.
15 L: You should drink some medicine.
16 A: Should I take my temperature?
17 L: Yes, you should.
18 A: Should I go swimming?
19 L: No, you shouldn't. You should wear...
20 A: "A jacket".
21 L: A jacket. Ετελειώσαμε; *[Are we done?]*
22 A: We already talked about these pictures. We're finished.
23 L: Ετελειώσαμε. *[We're done.]*

There were not many instances of scaffolded help in the conversations of Amy and Laura. The only two instances occurred on line 11 of S1 and on line 20 of S2, when the proficient student, Amy, provided the L2 phrase for Laura. The first instance occurred after Laura asked for help, but the second occurred when Amy realised that Laura was stuck.

Looking at the use of L1 metalanguage, Amy began the conversation on S1 by asking Laura which role she would like to play. This was an instance of *intrasentential* codeswitching (Myers-Scotton, 1997), as she began her question in Greek, but finished it in English. When Laura responded in English (line 2), Amy organised the task by telling her in the L1 to start the conversation. The next instance of L1 use in this conversation occurred on line 10, when Laura asked Amy how she could ask about taking the temperature in English. Amy responded by giving her the exact phrase she should use ("Get my temperature"). Lastly, the

girls signal the completion of the task by switching to Greek on lines 18 and 19 and stating that they had finished.

Moving on to S2, Laura began the conversation by instructing Amy in Greek to start asking the questions. Then, after Amy assigned the roles of the task in the L2, Laura expressed her satisfaction with her role by using the Greek equivalent of "good" (line 3). A few lines later, Amy switched to the L1 in order to direct Laura. She told her that she should ask the following question (line 12). Laura however, expressed her difficulty with formulating the question, while remaining in the L1. On line 21, Laura checked the progress of the task, when she asked in Greek whether the task was done. When Amy checked that they covered all pictures (line 22), Laura signalled the end of the task by switching to the L1. Having discussed their L1 use, it seems that L1 metalanguage served a number of purposes in Amy's and Laura's conversations, including managing the task, signalling the end of the conversation and for L2 equivalents.

As for the L2 metalanguage, the two conversations contained instances of different types of L2 metalanguage. On lines 1 and 2 of S1, Amy, who is the more proficient student, gave the opportunity to Laura to choose the role she would like to undertake. As shown in the transcript, Amy's question began in Greek but then switched to English. This is a type of creative L2 metalanguage, as no such question was given to this group on previous weeks. Laura responded in English, again with a phrase that was not previously presented to this group.

Having been given a handout with L2 phrases in S2, Amy chose to allocate the roles for this task, by using some of the newly introduced phrases. On line 2, her clarification that "Student A is the patient. Student B is the doctor" and her allocation of roles with the phrases "I want to be student A. You are the doctor" were all phrases that appeared on the handout. On line 3,

Laura showed her approval with the allocation of roles in Greek ("Good"). Then however, she continued in English with the phrase "Let's start", which did not appear as such on the handout of that week or on those of previous weeks. It was in fact a variation of other previously given phrases, such as "I start". This shows creativity on the part of an LP student.

Moving on to line 9, Laura used the phrase "What does this picture show?" from the handout of that week. On the following line, Amy gave a creative response in English describing the picture shown to her by Laura. Moreover, when on line 13 Laura said in Greek that she did not know how to describe the picture, Amy responded in English and tried to explain to Laura what she should say. Amy's response here falls into the creative L2 metalanguage category. The final instance of L2 metalanguage in the above transcript appeared on line 22. Amy checked the status of the task by borrowing a phrase from the handout of that week ("We already talked about these pictures"). Then she became more creative with her use of English again by commenting that they were finished. This latter phrase did not appear on the handout of that week or on those of previous weeks.

As with the previous example from this group, there are many instances in this example demonstrating that L2 metalanguage was used both under the guidance of the handouts, but also creatively. Creative L2 metalanguage did not only come from HP students, but from LP students as well. Both examples from the CS group therefore, suggest that the use of metalanguage in the L2 is possible. The fact that L2 metalanguage was used both in S2, as well as S1, could be used as evidence supporting the part of the hypothesis that predicts increased L2 practice due to the recycling of metalanguage into the L2.

ii. <u>*L2 metalanguage of EO group*</u>

This section investigates whether TR+ offers the opportunity for enriched output in terms of L2 metalanguage under the EO conditions. It should be noted here that some pairs in this group were not able to stick to the L2-only rule at all times, especially the LP students. Nevertheless, this did not affect fidelity to the conditions, as it was already shown in the Methodology chapter (Figure 3.5.1) that the EO group kept their L1 use to a minimum, in relation to the L1 use of the CS group. This section includes transcripts of pairs who managed to remain to the L2 for metalanguage the most at both time points. As with previous transcripts, commentaries on scaffolding, L1 metalanguage and accuracy feedback follow transcripts where appropriate.

The first example below presents the conversations of a HL proficiency pair, namely Colin and Paul, which took place on week 3. This was the week when they were asked to complete a narrative task based on a sequence of pictures. The underlined utterances present instances of L2 metalanguage.

Transcript 5.3.7 Example L2 metalanguage of EO group on week 3: Colin (HP) & Paul (LP)

Session 1

1 C: Ok. Πρέπει να πούμε την ιστορία. *[We have to tell the story]*
2 P: A.
3 C: Εγώ λαλώ το ένα, το δυο, το τρία. Εσύ το τέσσερα και το πέντε. *[I'll talk about picture one, two and three. You'll talk about four and five.]*
4 P: [inaudible]
5 C: Μα εν τζε ξέρουμε. Εν τζε γράφει μας. *[But we don't know. It doesn't say.]*
6 P: And then the way to circus.
7 C: Οι ετσι που ενι. *[No. It's like this.]*
8 P: Ok. Ξεκίνα πε τα τρία. *[You start. Talk about these three]*
9 C: Yesterday...
10 [I remind them of the L2-only rule.]
11 P: <u>Let's start</u>. Yesterday, Sally and Lil...
12 C: "And Tom."
13 P: Sally and Lil and Tom saw a poster of circus.
14 C: They went to circus. It was amazing. It was big. <u>Your turn.</u>
15 P: <u>We're in the third.</u>
16 C: "There was a poster"
17 P: There was a poster. Circus. It was writing. Circus. Ten dollars. Children five dollars.
18 C: <u>Ok my turn</u>. They went to cafe to eat.

19 P: Then two clowns, two f...fatty clowns, fat clowns, funny clowns get on the table and started dancing samba.

Session 2
1 C: Yesterday Tom and Lilly saw the poster of a circus.
2 P: Tom, Sally and Lilly decided to go to the circus.
3 C: Yesterday Tom, Sally and Lilly went to the circus.
4 P: They went to the circus by bus.
5 C: They went to the circus but they had no money. Your turn.
6 P: They didn't have money to pay the tickets.
7 C: So they went to the cafe and they drank coffee.

In terms of scaffolding, Colin, who was the more proficient student, offered his help to Tom twice. On lines 12 and 16 of S1, Colin helped Peter to either complete his utterances or start new ones by giving him the exact phrases that he should use. Regarding L1 use, the pair used some L1 metalanguage on S1. Despite the L2-only rule, Colin started the conversation in the L1 in order to state what they have to do. Then on line 3, Colin remained in the L1 when he attempted to organise the task by allocating three pictures to himself and two to Paul. With their discussion on lines 5 to 7, they continued discussing how to solve the task in Greek. Finally, on line 8 Paul directed Colin in Greek to start the narration. Finally, accuracy feedback seems to have been of some benefit to the students in this example. While in S1, both students omitted the articles "a" or "the" when referring to the circus or to the cafe (lines 13, 14 and 18), in S2 they used the two articles (lines 1, 3, 7) possibly because this was one of the errors that were included in the accuracy feedback. Other errors that were made on S1 (e.g. line 17, 19) did not appear in S2, not because they were not corrected, but because these phrases were omitted.

The rest of the metalanguage that took place in both sessions was in the L2. On line 9 of S1, Paul marked the beginning of their conversation with the phrase "Let's start". Then on line 13, having finished describing his picture, Colin reminded Paul in English that it is his turn. On the following line, Paul commented in English that they were on the third picture at that

point, attempting perhaps to check the progress of the task. Finally, Colin took his floor on line 17, when he said that it is his turn in English.

When the pair repeated the task, their use of metalanguage was minimised. The only instance of metalanguage was on line 5, where Colin gave Paul the floor by remaining in English. It is not surprising that the amount of metalanguage has been reduced from S1 to S2 however. Since the students were already familiar with what they should do in order to complete the task successfully, they started by using task language immediately. In terms of the inquiry of the present research question, which relates to enriched output, L2 metalanguage does not seem to evolve in any way when the task is repeated with accuracy feedback.

Another example of L2 metalanguage use in the EO group comes from the interactions between Miranda and Melanie; a HL proficiency pair. Their conversation took place in week 8, when they were asked to complete an information-gap task by giving instructions in order to find one another on a map. The transcripts below show the conversations that took place on the two sessions. The underlined phrases highlight L2 metalanguage.

Transcript 5.3.8 Example L2 metalanguage of EO group on week 8: Melanie (HP) & Miranda (LP)

Session 1

1 Mel: I will be student A.
2 Mir: I am student A.
3 Mel: Oh, yes.
4 Mir: Where are you?
5 Mel: I'm outside the cinema.
6 Mir: Ok. You have to go straight on. Turn right. Go past the park on your right. Where I am?
7 Mel: Wait. I will go straight. Turn right. You said turn right. You are at the supermarket?
8 Mir: Υποτίθεται εν supermarket. *[It's supposed to be a supermarket.]* No, I will say it again. Ok. You have to go straight on. Turn right. Go past the park on your right. Where I am? Oh sorry, sorry, sorry, sorry.
9 Mel: You are at the tennis?
10 Mir: At the tennis.
11 Mel: Court?
12 Mir: Uh, huh. I'm there.
13 Mel: You are at the tennis court. Where I am?
14 Mir: Now ο student A λαλεί το part two. *[Now student A talks about part two.]*
15 Mel: You are at the tennis court. Yes?
16 Mir: Yes.

17 Mel: You have to go straight to the park.
18 Mir: Yes.
19 Mel: Then, you...you have to turn right...and at the...at the first road you have to turn right and go straight and I am at the supermarket.
20 Mir: You are at the cinema. You said to me I have to go to the park. Then I have to turn to the right and then...
21 Mel: You have to...
22 Mir: Right, right, right.
23 Mel: Right.
24 Mir: Yes. And you said then I have...said it again please.
25 Mel: You have to go straight on the park.
26 Mir: Yes.
27 Mel: You have to turn right.
28 Mir: Yes, but this is the right. There are two rights.
29 Mel: What is the right?
30 Mir: Right. I writing with my right hand.
31 Mel: Yes but this is right.
32 Mir: This is left. Yes. I'm saying to you. You're saying to me go straight and then I am at the park...
33 Mel: A, sorry. Where are you?
34 Mir: I'm in the library.
35 Mel: You have to go right at…at the second road…at the second road you have to go straight down.
36 Mir: You're at the supermarket.
37 Mel: Yes.
38 Mir: Right. Μετά από τόσες ώρες εκαταφέραμε τα. *[After so many hours we made it.]*

Session 2

1 Mir: Student A είσαι. Άτε τέλειωνε. *[You are Student A. Go on.]*
2 Mel: Where are you?
3 Mir: I am outside...πού είμαι? *[Where am I?]* I am outside of the cinema.
4 Mel: Em. Ok. You have to go straight on, turn right, go pa...go past the park on our right. Where I am?
5 Mir: ?
6 Mel: To go straight on on right, go past the park on right, where I am?
7 Mir: Oh you are at the cinema. E sorry. You are at the….outside the tennis court. Now you are at the tennis court.
8 Mir: Where are you?
9 Mel: I'm outside the library.
10 Mir: The library?
11 Mel: Yes.
12 Mir: Library, library [looking for it on the map]. Oh yes!
13 Mel: Where I have to go?
14 Mir: Ee go to the park.
15 Mel: Yes.
16 Mir: Then when you are at the park you have to turn right.
17 Mel: Ok.
18 Mir: And then go to the…ok. You are now…
19 Mel: At the library.
20 Mir: Yes. Go straight on.
21 Mel: Yes. I have to go straight on.
22 Mir: Then turn right and go straight on. And you will find me where I am.
23 Mel: No.
24 Mir: You are not clever. [laughs]
25 Mel: I have to…I have to…I have to turn right at the first road or at the second road?
26 Mir: Yes.
27 Mel: At the second?
28 Mir: One minute one minute. Go to the park.

29 Mel: Ok.
30 Mir: Go right.
31 Mel: Yes but the second? The first or the second road?
32 Mir: The second.

While there are no instances of scaffolding in the above transcripts, a few instances of L1 metalanguage occurred. The first instance was on line 14 of S1 and it was a case of intrasentential codeswitching (Myers-Scotton, 2006). Miranda began the sentence in English ("Now"), but then continued in Greek. Her utterance signalled the beginning of the second part of the task. The second and final instance of codeswitching in their conversation was again expressed by Miranda at the very end of the conversation (line 38). Having completed the task, Miranda commented on their performance by referring to the length of time it took them to finish the task in Greek (line 38).

Looking at the transcript of S2, again there were only a couple of instances of L1 use. The first instance occurred at the very beginning of their conversation, when Miranda reminded Melanie of her role as Student A, in order to remind her that she should start the conversation. The second instance of codeswitching occurred on line 3, when Miranda asked herself where she was, perhaps in order to buy some time to find where she was on the map. This is considered to be an instance of *private speech* (Antón & DiCamilla, 1999), since it is obvious that Miranda did not expect a response from her interlocutor. Having gone through all L1 instances in the two transcripts, it is worth noting that all these instances were initiated by Miranda, who was the LP student. It seems safe to argue therefore, that it is much more difficult for LP students to follow the L2-only rule, than for HP students. Considering accuracy feedback, no erroneous utterances were spotted in their conversation during S1, so no claims can be made in terms of the benefits of accuracy feedback.

The pair used metalanguage in the L2 as well. On the first three lines of S1, the pair assigned roles using the L2. Then, when Miranda gave Melanie the instructions, Melanie ensured that

she had followed them correctly on line 7. She asked Miranda in English to pause ("Wait") so that she could repeat what Miranda said. When they realised that something went wrong with the instructions, Miranda decided to repeat her instructions (line 8), again expressing herself in the L2. Later on in the conversation, when they switched roles, Miranda checked with Melanie that she followed her instructions correctly by repeating what Melanie had said (line 20). On line 24, Miranda asked Melanie in English to repeat what she had said. Due to all this repetition of instructions, it is not surprising that L2 metalanguage in this example is somewhat repetitive. Finally, on lines 28 to 33, the girls stick to English while they have an argument about which direction is right and which is left.

On S2, the pair used metalanguage at fewer instances. As mentioned earlier, this is not surprising because they were already familiar with the content of the task. Metalanguage was mainly used from lines 24 to 28. On line 24, when Melanie failed to successfully follow Miranda's instructions, Miranda joked in the L2 that Melanie was not clever. Then on line 25, Melanie attempted to understand what went wrong, when a few lines later, Miranda took the floor by asking Melanie to pause (line 28). Although all this took place in the L2, L2 metalanguage did not seem to change in any way or to become more creative when the task was repeated.

As a conclusion to the L2 metalanguage section, it seems that the conditions have a role to play in the evolution of L2 metalanguage. While the transcripts of the CS group include examples of L2 metalanguage becoming more creative with TR+, no such examples exist in the transcripts of the EO group. An explanation for this difference could be found in the conditions of the experiment. The CS group received feedback on metalanguage and, as discussed in this section, they used it either as it was, or they experimented with it on subsequent occasions. On the other hand, the EO group did not have this foundation of L2 metalanguage to build on, when they repeated tasks. It could be argued, therefore, that by

receiving feedback on L2 metalanguage and by repeating tasks, the CS group produced more enriched output than the EO group.

5.4. Conclusions

The main focus of this chapter was to examine whether the content of the output changed in any way when tasks were repeated with feedback. More specifically, the qualitative analysis focused on changes of task language or L2 metalanguage. Changes in the sense of expanding ideas, adding new ideas or using a larger range of vocabulary, were considered as evidence of 'enriched output'. These changes were illustrated with transcripts from both the CS and the EO group. The transcripts were also analysed in terms of three other aspects which were considered to be relevant. These are scaffolded help, the use of L1 metalanguage and the effects of accuracy feedback. This section summarises the conclusions reached for each one of these three aspects first and then reviews the conclusions on enriched language.

5.4.1. Scaffolding

Having reviewed the transcripts, it became clear that the students often provided help to one another in order to complete the tasks successfully. Their help was often *direct*, that is when the one student was telling, or rather whispering to, the other student exactly what s/he should say. This kind of help often occurred when one of the students was of a higher proficiency than the other. At times, in an attempt to support their interlocutors, HP students took over the tasks and guided their LP partners. Direct help was also evident when HP students identified errors in the output of their LP counterparts, and immediately corrected them by repeating the phrase error-free. In addition to direct forms of help, scaffolded help occurred

sometimes in more *indirect* forms; for example, when HP students signalled an error and gave the opportunity to LP students to reconsider their own output.

All the above provide evidence for the existence of a collaborative environment. The students worked together in order to solve the tasks and achieve the best possible result.

5.4.2. L1 metalanguage

A number of L1 metalanguage instances were identified in the transcripts. The use of L1 metalanguage seems to serve certain functions which are common in both conditions. Metalanguage expressed in the L1 was mainly used for assigning roles in a task or deciding who is going to start, when the roles were pre-set. Another very common reason for using L1 metalanguage was for achieving a shared understanding, or else *intersubjectivity,* on what they have to do (Antón & DiCamilla, 1999). The students often re-stated the goals of the task and tried to work out what they have to do in order to complete it successfully. Moreover, the L1 was used for finding the L2 equivalents of words. Also, it was often used in order to signal the beginning or the end of the task or a part of the task.

Students who needed some help often switched to Greek to ask for it, while students who provided help often directed their interlocutors in the L1. Furthermore, the L1 was used in order to comment on their own performance. In addition, when students wanted to check the progress of the task often switched to the L1. Finally, Greek was used to externalise *inner speech*; a term encountered in the literature for instances where speakers vocalise their private speech, not because they seek their interlocutor's response, but because they want to re-direct their thinking.

It is fair to say therefore that L1 metalanguage served a number of different purposes and probably played an important part in solving the tasks.

5.4.3. Effects of accuracy feedback

The question of whether accuracy feedback was implemented by the students when tasks were repeated was investigated. On the whole it seems that students were not able to instantly incorporate the feedback they received. While there are some examples of students correcting parts of their output based on the feedback, this evidence is not enough. This could suggest that learners in this study were not able to absorb this input instantly and incorporate it into the task. This should be taken into consideration when interpreting the results of the quantitative analysis on accuracy.

5.4.4. Enriched output

This part provides a response to RQ2. As discussed in this chapter, enriched output comes in the form of two types of language in the present study: task language and L2 metalanguage.

i. Task language

Looking at the results of task language first, there is evidence in the data suggesting that TR+ may either have a positive or a negative effect on task language. There are cases where students of both groups took the opportunity of repeating the same task and used it in order to experiment with their output and produce new ideas. However, this was not always the case, as there are examples in the data suggesting the exact opposite; that repetition of the same task does not always enrich the content of task language, but rather restricts it. Two reasons

of why this is happening were proposed: first, learners often lose interest in what they are doing when they complete the same activity; second, learners do not feel challenged by the level of the task and do not wish to experiment with task language. Nonetheless, the first reason was considered to be more plausible for the case of the present study because the students' loss of interest and frustration was also documented in my log-book during the intervention, as well as in the teachers' interviews discussed in the previous chapter. Although these students were given feedback on the second occasion, the fact that the task they had to complete was the same seemed to have had an impact on their performance at times. In response to RQ2 therefore, the results cannot support that TR+ always leads to enriched task language.

ii. <u>*L2 metalanguage*</u>

Before discussing the results on how L2 metalanguage evolved, it would be interesting to note that a number of functions served by L2 metalanguage were similar to those served by L1 metalanguage, just like the findings of DiCamilla and Antón (2012) without the effect of the age factor. As with L1 metalanguage, learners used L2 metalanguage in order to assign roles or determine who would start the interaction, with phrases such as "I start". In addition, L2 metalanguage was used for signalling the beginning or the end of the task or a part of the task, just like L1 metalanguage. Phrases like "We start now" or "Now we finish" were found in the data. Moreover, intersubjectivity was often reached with L2 metalanguage (e.g. "One picture for you and one picture for me"). Finally, as with L1 metalanguage, L2 metalanguage was used for checking the status of the task (e.g. "We already talked about these pictures).

Other functions, however, were only served by L1 metalanguage. These include commenting on their performance, externalising inner speech and asking for help. What can be concluded

from this discussion on functions therefore, is that L2 metalanguage can often serve the same functions as L1 metalanguage, but L1 metalanguage seems to serve a greater range of functions. Perhaps in time, as proficiency progresses, these students will be able to use the L2 for whole range of functions served by the L1. As it was suggested in the DiCamilla and Antón (2012) study, more advanced learners used the L2 for functions that beginner learners used their L1 for. Nevertheless, the fact that most functions could be served by both languages seems to provide support to part of the hypothesis suggesting that there is no indication that L2 metalanguage cannot provide the same cognitive support as L1 metalanguage when solving tasks.

Moving on to the evolution of L2 metalanguage, which is what RQ2 is concerned with, the results on L2 metalanguage differed according to the two experimental conditions. The data coming from the CS group demonstrated that L2 metalanguage was often enriched when tasks were repeated. Students did not just use the feedback from the handouts of that week or of previous weeks, but they demonstrated that they were trying to be creative with it. In addition, some students in this group started building on this feedback and being more creative with their use of L2 metalanguage. There are utterances in both sessions of certain weeks that represent purely creative L2 metalanguage. It has been argued therefore that this provides evidence that learners in this context are able to use the L2 for metalanguage while completing group tasks and that this idea is not as far-fetched as the teachers reported it to be in their interviews. This creative L2 metalanguage was not only used by HP students, but by LP students as well. Practising this kind of language on a regular basis therefore, increases the opportunities for learners to speak the L2, while it meets the overall aims of task-based learning, which is to communicate effectively by using authentic language.

As far as the EO group is concerned however, there is no evidence of positive effects of TR+ on their L2 metalanguage. In fact, the L2 metalanguage of the EO group was often reduced

on S2. While on the one hand, one could argue that this is perhaps due to the fact that learners were already familiar with the purposes of the task and did not need to discuss how to complete it, a counter-argument could point at the CS group, who was experimenting with L2 metalanguage, despite the fact that they were also already familiar with the task. The explanation proposed in the chapter as the reason for this 'undeveloped' metalanguage on S2, was the lack of feedback on such language. This group could not expand their L2 metalanguage simply because they did not have the required foundation to do so.

Concluding this section therefore, it seems that the feedback component of the present study had an important effect on the performance of the repeated task. By providing recycled L2 metalanguage, the CS group had the opportunity to use it as foundation for further experimentation and creativity. It appeared therefore that the output of the CS group was enriched when the tasks were repeated partly because their L2 metalanguage evolved in some way. On the contrary, the performance of the EO group did not resemble that of the CS group because they did not receive the same foundation to experiment. In their case therefore, it appeared that their use of L2 metalanguge either remained as it was or it was reduced. In response to RQ2 thus, it seems that TR+ can lead to advanced L2 metalanguage, when relevant feedback is provided.

5.5. Chapter summary

This chapter presented the qualitative analysis conducted for the data coming from the recordings of student interactions during the intervention. The purpose of the analysis was to respond to the second research question which inquired about the potential enrichment of the content of the output when a task was repeated with feedback. Two forms of output were examined: task language and L2 metalanguage. For each one of these forms of output,

transcripts with the conversations taking place on both sessions were provided, by taking into consideration the conditions under which the pairs operated. Apart from the main focus of the chapter which was the evolution of task language and L2 metalanguage, the transcripts were analysed in terms of three other aspects: scaffolded help, the role of L1 metalanguage and accuracy feedback.

Interesting results included the fact that L1 metalanguage and L2 metalanguage seemed to have served a number of similar functions, providing support to the argument of the hypothesis suggesting that L2 metalanguage can provide the same cognitive support as L1 metalanguage, when completing tasks. In terms of accuracy feedback, it was found that it was difficult for the EO group to incorporate the feedback given to them when they repeated tasks. Very few errors were corrected due to accuracy feedback.

In response to RQ2, the analysis did not provide definite results in terms of task language. There was evidence suggesting both increased creativity and reduced creativity of task language after TR+. In terms of L2 metalanguage however, the conditions seemed to have played an important role. While there was no evidence suggesting that TR+ favoured L2 metalanguge under EO conditions, there was evidence suggesting that CS conditions helped with the creativity of L2 metalanguage. One explanation proposed for this result was that the CS group demonstrated enriched L2 metalanguage because they were able to experiment with the feedback on L2 metalanguage they received. What should be added here is that the creative L2 metalanguage used by the CS group suggested that the idea of learners using L2 metalanguage in this context was not far-fetched. These results are discussed further in relation to existing results from the literature in the Discussion chapter.

CHAPTER 6: Quantitative Results

6.1. Introduction

This chapter presents the results of the quantitative analysis. It provides a response to the last three research questions, which inquire about the effects of the intervention. The data come from the three oral production tests (pre & post-tests). Prior to the research questions, the results of four possible confounding variables are discussed. Based on these variables, an attempt to match the participants of the three groups is made. The chapter therefore consists of four main parts. The first part presents the results of the possible confounding variables, while the remaining three discuss the results of the three research questions respectively.

6.2. Possible confounding variables

The four possible confounding variables are *L2 proficiency, productive vocabulary, socio-economic status* and *prior L2 contact.*

6.2.1 Second Language Proficiency

The proficiency level of the students was determined by the scores of the C-test. As these will be tested for between-group differences, the assumptions of normally distributed data and of homogeneity of variance were considered. A graph with the distribution of the data is included in Appendix 35. In order to explore these assumptions further however, relevant statistical tests were employed. The Kormogorov-Smirnoff test (thereafter the K-S test) indicated that the proficiency scores were significantly non-normal, D (75) = .21, $p < .001$, while the Levene's test indicated that the variances were significantly different in the three groups, F (2, 72) = 3.14, $p = .049$. The boxplots below show the range of scores obtained by each group. Table 6.2.1 that follows presents the exact figures of these scores.

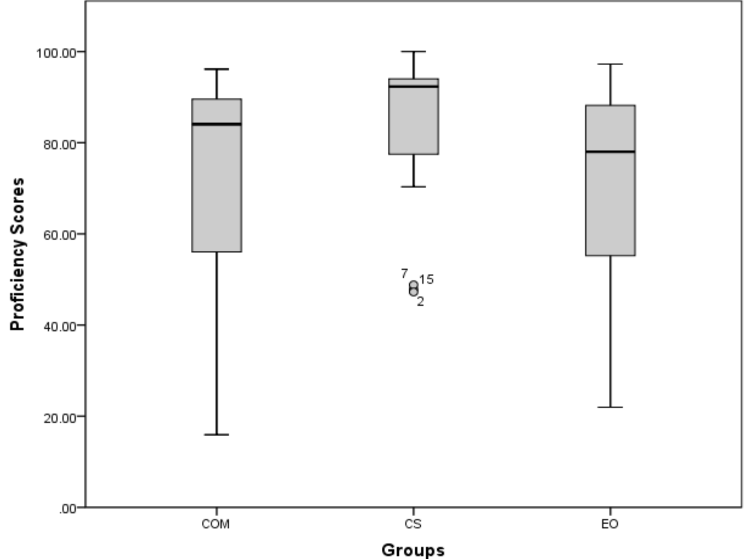


Figure 6.2.1 Boxplots: L2 Proficiency scores of the three groups (C-test scores)

Table 6.2.1 Descriptive statistics: L2 Proficiency (C-test)

Group	N	Mean (SD)	Median	Range	Min	Max
Codeswitching	21	83.12 (16.88)	92.31	52.75	47.25	100.00
English-only	35	69.53 (23.71)	78.02	75.27	21.98	97.25
Comparison	19	71.77 (25.85)	84.06	80.22	15.93	96.15

As the assumptions of normality and homogeneity of variance were not met, the Kruskal-Wallis test, which is the non-parametric equivalent test of a one-way ANOVA, was chosen to explore the differences between the three groups. This showed that proficiency scores were significantly affected by groups, H (2) = 5.97, p = .051. Mann-Whitney tests were used to follow up this finding. As a three-way comparison was conducted (CS-EO, CS-COM, EO-COM), a Bonferroni correction was applied so all effects are reported at a .0167 level of significance. The difference between the CS and the COM group was non-significant, U = 140, $p > .0167$; the difference between the EO and the COM group was non-significant, U = 305.5, $p > .0167$; but the difference between the CS and the EO group was significant, U = 226.5, $p < .0167$, r = - .32.

6.2.2 Second Language Productive Vocabulary

As described in the previous chapter, productive vocabulary was measured with a word translation test. The distribution graph for the data is also included in Appendix 35. The K-S test showed that the productive vocabulary scores were significantly non-normal, $D\ (75) = .15$, $p < .001$, while the Levene's test indicated that the variances were not significantly different in the three groups, $F\ (2, 72) = 1.07$, $p = .35$.

Figure 6.2.2 below presents boxplots, which represent the spread of the scores of this test, while table 6.2.2 gives the descriptive statistics.

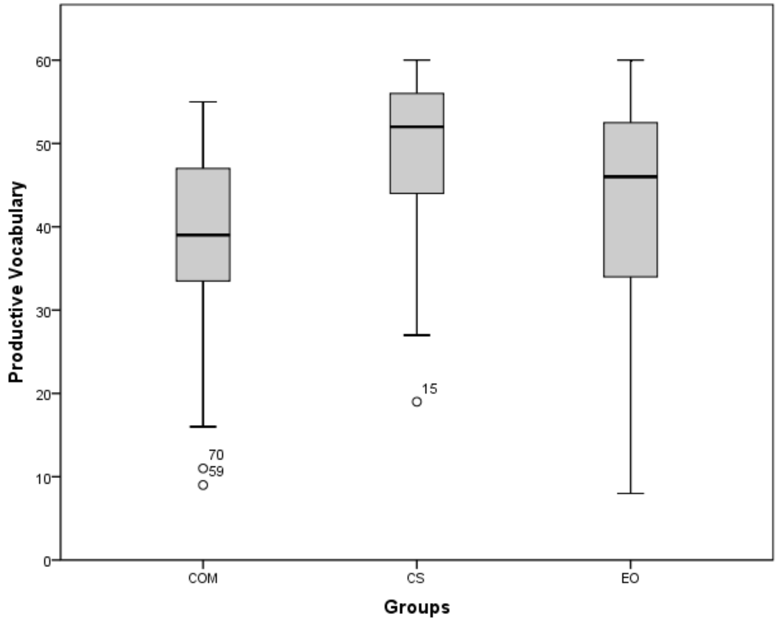


Figure 6.2.2 Boxplots: Productive Vocabulary Scores of the three groups (Translation test)

Table 6.2.2 Descriptive Statistics: Productive Vocabulary Scores

Group	N	Mean (SD)	Median	Range	Min	Max
Codeswitching	21	48.95 (10.67)	52	41	19	60
English-only	35	42.03 (13.25)	46	52	8	60
Comparison	19	37 (13.89)	39	46	9	55

As the assumption of normality was not met, the Kruskal-Wallis test was used to explore the differences between the three groups. This showed that the productive vocabulary scores were significantly affected by groups, H (2) = 10.30, $p < .01$. A three-way comparison was conducted with the help of Mann-Whitney tests. A Bonferroni correction was applied so all effects are reported at a .0167 level of significance. The difference between the CS and the COM group was significant, $U = 83$, $p < .0167$, $r = -.5$; the difference between the EO and the COM group was non-significant, $U = 253$, $p > .0167$; and the difference between the CS and the EO group was non-significant, $U = 332.5$, $p > .0167$.

It is worth mentioning that the productive vocabulary scores were significantly correlated with C-test scores, $r = .71$, $p < .01$. This suggests that the students who did well on the C-test, did also well on the productive vocabulary test. The high correlation validates the use of the C-test as a proficiency measure.

6.2.3 Socio-economic status

Each student received a score for SES based on the answers given by their parents on the SES questionnaire. The distribution graph for these scores appears in Appendix 35. According to the K-S test, the SES scores were significantly non-normal, D (75) = .11, $p < .05$, while the Levene's test indicated that the differences between the variances of the groups were non-significant, F (2, 72) = .76, $p = .47$. The boxplots below show how these scores are spread for each group.

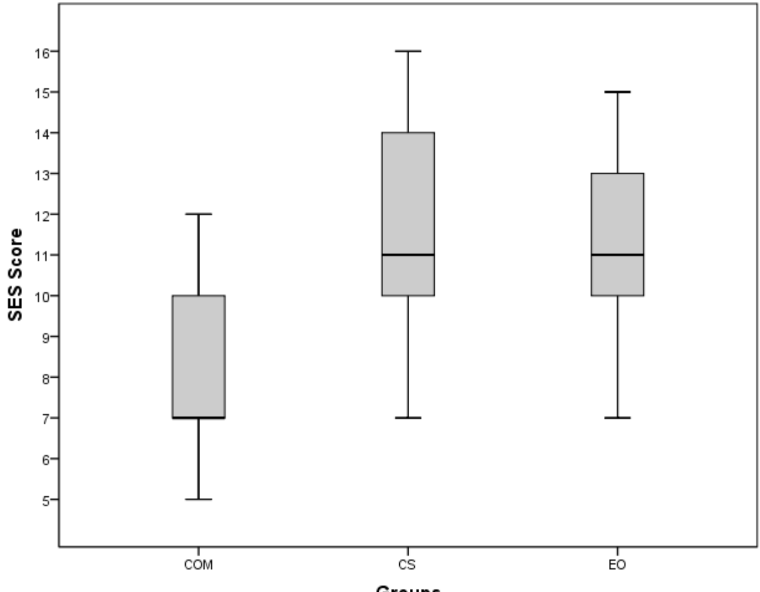


Figure 6.2.3 Boxplots: SES scores of the three groups

Table 6.2.3 Descriptive Statistics: SES scores

Group	N	Mean (SD)	Median	Range	Min	Max
Codeswitching	21	11.62 (2.40)	11	9	7	16
English-only	35	11.17 (2.49)	11	8	7	15
Comparison	19	8.05 (2.07)	7	7	5	12

As the assumption of normality was not met, the Kruskal-Wallis test was conducted. This showed that SES scores were significantly different between the groups, H (2) = 20.06, $p < .001$. As a three-way comparison of Mann-Whitney tests was conducted, the Bonferroni correction was applied again at a .0167 level of significance. The difference between the CS and the COM group was significant, $U = 55$, $p < .001$, $r = -.62$; the difference between the EO and the COM group was significant, $U = 115$, $p < .001$, $r = -.54$; and the difference between the CS and the EO group was non-significant, $U = 332.5$, $p > .0167$.

The students were allocated to a *High SES* group or a *Low SES* group based on a median split. The histogram below shows this classification graphically.

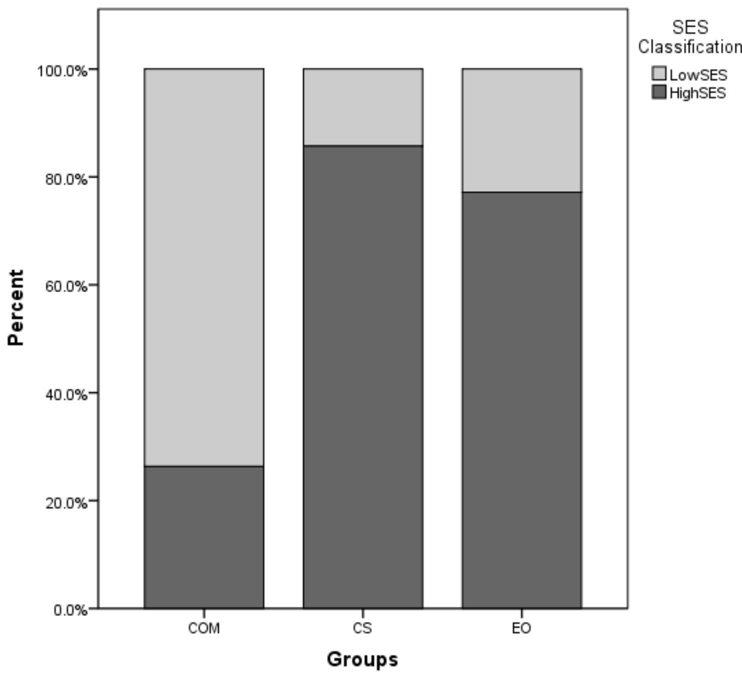


Figure 6.2.4 Histogram: SES classification of the three groups

Based on the figure above, it is clear that in general the COM group is of a lower SES in comparison to the other two groups. This discrepancy could be explained by the fact that the school representing this group was not situated as close as the other two schools. Nevertheless, this difference in terms of SES should be kept into consideration, when interpreting the results of this study.

6.2.4 Prior L2 contact

The *prior L2 contact* variable is an ordinal variable, which ranks the contact the participants had with the English language prior to the intervention. The data for this variable come from the responses given by the participants on a relevant questionnaire.

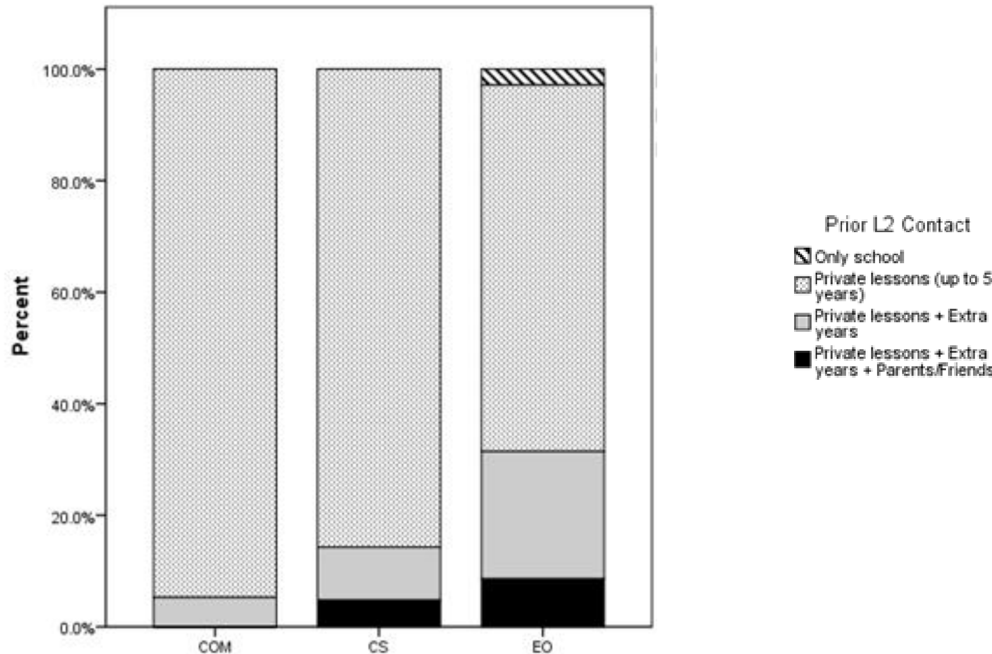


Figure 6.2.5 Histograms: Prior L2 contact classification

It is clear from the figure above that the majority of the students in all groups had been taking private lessons of English for up to five years. A small percentage in all groups had received private English lessons for more than five years. Students, who have English-speaking parents or close friends, represent a very small percentage in the CS and EO group, while the EO group has a very small percentage of students who received English lessons at school only. From this classification, it could be said that the CS and the EO group are more likely to have participants who are more familiar with the English language, than the COM group.

Overall, the significant differences found are: proficiency difference between the CS and the EO group; a productive vocabulary difference between the CS and the COM; and an SES difference between the experimental groups and the comparison group. The differences

should be kept into consideration while discussing the results of this chapter. These differences were dealt with in the matching process described below.

6.2.5 Matching Process Results

An effort was made to match the three groups in order to make them as comparable as possible. The matching process was based on the four variables discussed above. It is expected that the matching process will not achieve perfect matching for a number of reasons. As Dörnyei (2007) argues, perfect matching is never possible. This is because in social science research numerous confounding variables may affect the situation. Moreover, the uneven size of the groups makes it more difficult to find a match for every student in every group. The results of the matching process, therefore, are reported in terms of the number of variables that are matched. This is better explained in Table 6.2.4 below.

Table 6.2.4 Matching process results

	Perfect match (4 variables matched)	**Nearly perfect match** (3 variables matched)	**Close match** (2 variables matched)	**Total number of students considered**
Number of students that match across 3 groups	6 (2 from each group)	33 (11 from each group)	12 (4 from each group)	51
Number of students that match across 2 groups (CS & EO group)	4 (2 from each group)	4 (2 from each group)	0	8
Number of unmatched students				16
				75

Perfect matching, which is a match of all four variables across all three groups, was only achieved with two students. When using three variables, that is a *nearly perfect match*, eleven more students were matched across all groups. From the remaining students, four were matched by two variables across all groups. This was the best possible match that could be achieved across the three groups. In order to continue the process with the remaining participants, matching students across two groups, namely the CS and the EO group, was attempted. Perfect matching was successful with two students. Two more students were nearly perfectly matched. From the whole sample, sixteen students remained without a match. This process showed that the three groups are comparable to some extent.

6.3. RQ3: a) Does task repetition plus feedback lead to more willingness to communicate than no task repetition?

b) Does task repetition with codeswitching *plus* lead to more willingness to communicate than task repetition with English-only *plus*?

The third research question focuses on the concept of willingness to communicate (WTC). First, it examines whether TR+ leads to more WTC than no task repetition. Second, it investigates whether pedagogical package A leads to more WTC than pedagogical package B, as the hypothesis predicted. Two variables had been chosen as measures of WTC: *Speaking time* and *L2 syllables*.

6.3.1. Speaking time

As discussed in the Methodology chapter, speaking time was measured in seconds and it consisted of the time spent by each student speaking or pausing during the oral production tests (both question/answer & presentation part).

Graphs showing the distribution of the data for each time point appear in Appendix 36. Although the distributions look satisfactorily normal, the K-S test indicated that these were normally distributed only at time 1: pre-test, D (72) = .08, p = .20; post-test 1, D (72) = .13, p < .01; post-test 2, D (72) = .12, p < .01. As for the homogeneity of variance, the Levene's test indicated that this assumption had been met for all three time points: pre-test, F (2, 69) = .40, p = .67; post-test 1, F (2, 69) = 2.33, p = .10; post-test 2, F (2, 69) = .77, p =. 47. As the assumption of normality was not met, the between-group differences of speaking time were investigated with both the Kruskal-Wallis test and its parametric equivalent, namely the one-way ANOVA (see Appendix 37). The comparison revealed that the results of both types of tests were the same.

The most suitable test for the investigation of these data was the *mixed between-within ANOVA* (thereafter mixed ANOVA), which considers the interaction effect of the group variable and the time variable. As there was no non-parametric test equivalent to the mixed ANOVA, and since the comparison of the two types of tests shows no differences, the parametric mixed ANOVA was conducted. The mean scores of each group at each time point are shown in the line graph below (Figure 6.3.1), for which the descriptive statistics are shown in Table 6.3.1.

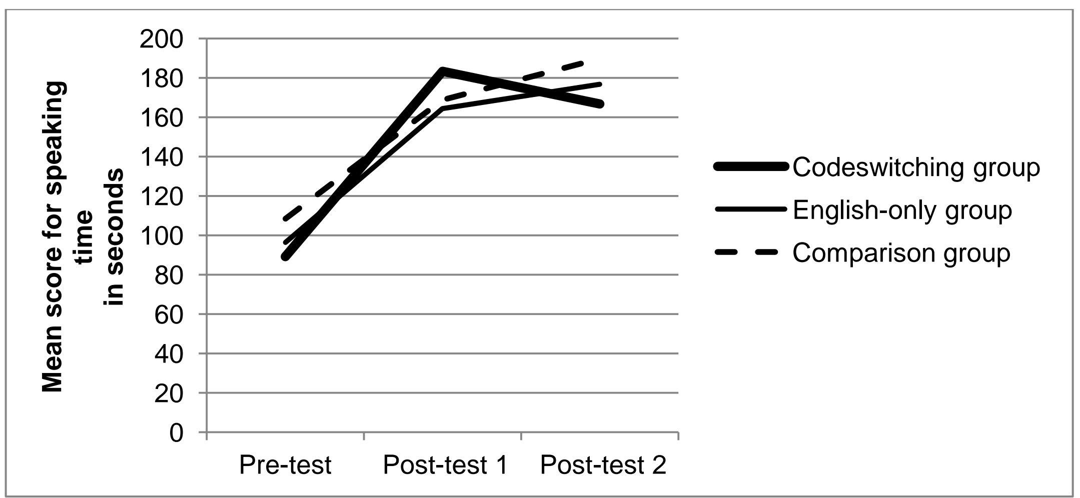


Figure 6.3.1 Line graph: Mean performance of speaking time of the three groups across time

Table 6.3.1 Descriptive Statistics: Speaking time

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	89.19 (33.52)	132.20	22.20
	EO	34	96.38 (38.72)	167	203
	COM	19	108.44 (36.85)	116.90	167.10
Post-test 1 (Time 2)	**CS**	21	183.21 (87.58)	385.30	479.80
	EO	34	166.34 (48.38)	171.80	256.10
	COM	18	168.67 (56.20)	207.20	300.80
Post-test 2 (Time 3)	**CS**	21	166.76 (73.33)	281	373.40
	EO	35	176.64 (57.48)	235.90	315.30
	COM	19	189.65 (90.52)	340.80	422.70

As the one-way ANOVA showed that there was no significant difference between the groups at pre-test for speaking time, F (2, 73) = 1.39, p = .26, the mixed ANOVA was conducted. As far as the within-group effect is concerned, the Mauchly's test showed the assumption of sphericity was met, χ^2 (2) = .52, p = .77. A significant main effect of time was found, F (2, 138) = 71.97, $p < .001$, *partial* η^2 = .51. According to Cohen's (1988, 1992) suggestions on what is a small or large effect, which are used for interpreting effect sizes throughout this chapter, this effect was very large as it explained 51% of the total variance of the model. More specifically, the result suggested that the time of the test significantly affected speaking time. Contrasts revealed that the difference lay between time 1 and time 2, F (1, 69) = 103.62, $p < .001$, *partial* η^2 = .60, a large effect that explained 60% of the variance, but not between time 2 and time 3, F (1, 69) = .07, p = .79. Pairwise comparisons indicated that the difference was also significant between time 1 and time 3, $p < .001$.

As for the between-group differences, a non-significant effect was found, F (2, 69) = .11, p = .90. This suggests that the overall level of speaking time did not differ according to groups. In

other words, if the time variable is ignored, there is no significant difference between the mean scores of the three groups. A non-significant interaction effect was found between the time of the test and the groups, F (4, 138) = 1.17, p = .33. This indicated that the change of speaking time across time did not differ according to groups.

6.3.2. L2 Syllables

The second variable measuring WTC is L2 syllables, which is the total number of L2 syllables produced by the student in the oral production tests. The distribution of the data at each time point is shown in Appendix 36. According to the K-S test, the assumption of normality was not met for the L2 syllables variable at pre-test, D (72) = .12, $p < .05$, post-test1, D (72) = .19, $p < .001$, or at post-test 2, D (72) = .11, $p < .05$. As for the homogeneity of variance, the Levene's test indicated that the assumption was met at pre-test, F (2, 69) = .98, p = .38, and post-test 2, F (2, 69) = 1.29, p = .28, but not at post-test 1, F (2, 69) = 6.76, $p < .01$. Due to these violations of normality and homogeneity of variance, the results of non-parametric tests were compared to those of parametric tests on these variables (Appendix 37). The comparison revealed that both types of tests uncovered significant differences between the three groups at all three time points. Since the results were the same therefore, it was decided to proceed with parametric tests. The line graph in Figure 6.3.2 below shows the mean scores of each group, while Table 6.3.2 presents the descriptive statistics.

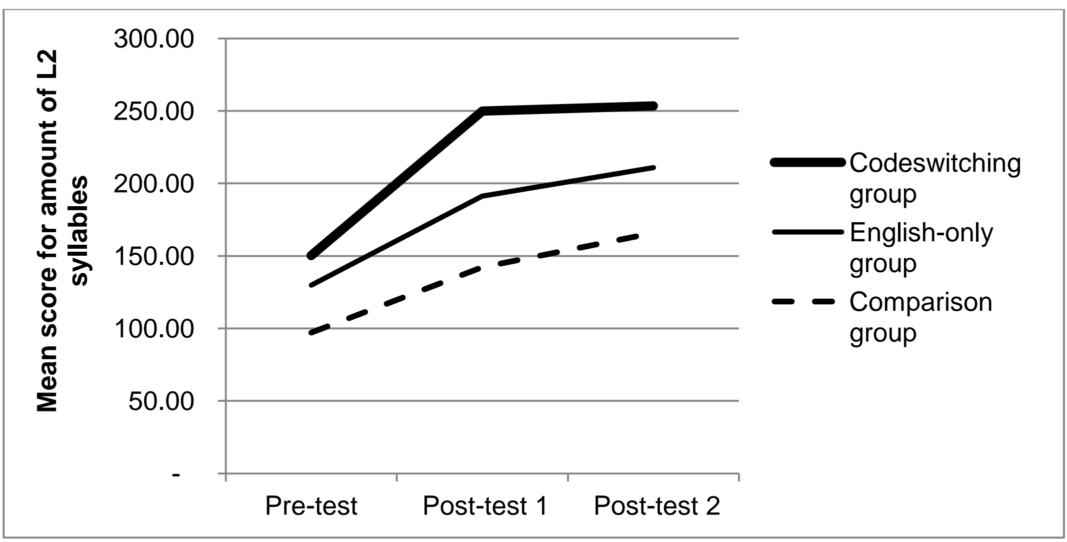


Figure 6.3.2 Line graph: Mean performance of the amount of L2 syllables of the three groups across time.

Table 6.3.2 Descriptive Statistics: L2 syllables

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	150.24 (73.43)	306	341
	EO	35	129.91 (71.66)	271	305
	COM	19	97.05 (52.30)	179	213
Post-test 1 (Time 2)	**CS**	21	249.86 (140.85)	514	599
	EO	34	191.21 (89.03)	427	487
	COM	18	142.11 (43.29)	175	247
Post-test 2 (Time 3)	**CS**	21	253.43 (123.72)	475	564
	EO	35	210.80 (89.25)	381	448
	COM	19	165.84 (86.71)	312	374

As with the speaking time variable, the mixed ANOVA test was the ideal test to examine the interaction of the group and the time variable. However, a one-way ANOVA at pre-test revealed that there was a significant difference for L2 syllables between the three groups, F (2, 74) = 3.11, $p = .051$, $r = .28$. This was a medium effect as it explained only 7.84% of the total variance. According to Hochberg's GT2 test, the difference laid between the CS and the COM group only, $p < .05$. Since there was a significant difference at pre-test, two ANCOVAs

were conducted, in order to look at the between-group differences at time 2 and time 3, with time 1 as the covariate.

The first ANCOVA revealed that the covariate, namely pre-test scores, was significantly related to post-test 1 scores, F (1, 69) = 79.21, $p < .001$, *partial* $\eta^2 = .53$. This was a very large effect as it explained 53% of the total variance. Having controlled for the effect of pre-test scores, a significant effect of groups was also found on post-test 1 scores, F (2, 69) = 3.20, $p < .05$, *partial* $\eta^2 = .08$. This effect was medium as it only explained 8% of the variance. Planned contrasts revealed a significant difference between the CS and the EO group, $p < .05$, and the CS and the COM group, $p <.05$, but not between the EO and the COM group, $p = .46$. According to the second ANCOVA, pre-test scores were significantly related to post-test 2 scores as well, F (1, 71) = 112.74, $p < .001$, *partial* $\eta^2 = .61$. The effect of these scores was very large as it explained 61% of the total variance. However, no significant effect of groups was found on post-test 2 scores, after controlling for the effect of pre-test scores, F (2, 71) = .95, $p = .39$.

6.3.3 Response to RQ3

The third research question compares the potential effects of TR+ on WTC, in relation to no task repetition. Then it tests whether pedagogical package A, that is task repetition with codeswitching plus feedback, leads to more WTC than pedagogical package B, that is task repetition with English-only plus feedback. TR+ is tested here by the two experimental groups, while task repetition only is represented by the comparison group. Table 6.3.3 below shows the results of the two measures of WTC, namely speaking time and L2 syllables.

Table 6.3.3 Results of WTC measures

WTC MEASURES					
Speaking time	Mixed ANOVA		**L2 syllables**	2 ANCOVAs	
	Time	*sig**** 1 & 2*** 1 & 3***		Pre-test	*sig** CS & COM*
	Groups	*ns*		Post-test 1	*sig** CS & EO* CS & COM*
	Time*Groups	*ns*		Post-test 2	*ns*

sig = significant; *ns* = non-significant; $p < .05$*; $p < .01$**; $p < .001$***; appr. = approaching significance

i. *Task repetition plus vs. no task repetition*

In response to the first inquiry, the data do not provide evidence that TR+ leads to more WTC, than no task repetition. The results of speaking time show a significant change across time, but the absence of a significant interaction effect suggests that this change across time does not differ according to groups. In terms of L2 syllables, the significant difference between the CS and the COM group partly supports that TR+ leads to more WTC than no task repetition. However, since a significant difference between the EO and the COM group was missing, no claims favouring TR+ can be made.

ii. *Task repetition with CS plus vs. task repetition with EO plus*

The second inquiry of the research question concerned the effects of the two pedagogical packages on WTC. A significant difference was found between the CS and the EO group in terms of L2 syllables only. Based on this variable only therefore, it could be argued that task repetition with CS plus leads to more WTC than task repetition with EO plus.

Despite the two inquiries, the fact that the CS group produced significantly more L2 syllables than the other two groups should not be disregarded. It is an indication that this group ended up being more willing to communicate than the others.

6.4. RQ4: a) Does task repetition plus feedback lead to more fluency than no task repetition?

b) Does task repetition with codeswitching *plus* lead to more fluency than task repetition with English-only *plus*, with no detrimental effects on accuracy?

This question looks at the effects of each condition on the development of fluency. First, it investigates whether TR+ favours fluency. Then it focuses on whether task repetition with codeswitching *plus* is more beneficial for fluency than task repetition with English-only *plus*. The inquiry assumes however, that any potential fluency development will come with no detrimental effects on accuracy.

As mentioned in the Methodology chapter, the question is answered with the analysis of three variable constructs: the *fluency variables,* which are speech rate, mean length of runs, pause rate and the dysfluency markers (repetitions, reformulations, incomplete words); the *content variables,* which consist of the verb ratio, L1 syllables and idea units; and the *accuracy variables,* which are the percentage of error-free clauses and the percentage of correct verb forms. In case there is a change in fluency, these variables will not only investigate whether accuracy is deteriorated, but whether the quality of content is affected as well.

6.4.1 Fluency Variables

i. Speech Rate

The first variable in the fluency variables construct is *speech rate,* that is syllables per second. The normality of the speech rate data for each time point is shown in graphs included in Appendix 38. These distributions seem satisfactorily normal. The K-S test confirmed that this is the case: pre-test, D (72) = .07, p = .20; post-test 1, D (72) = .06, p = .20; post-test 2, D (72) = .07, p = .20. In addition, the Levene's test confirmed that the data were homogenous at

all three time points: pre-test, F (2, 69) = .89, p = .41; post-test 1, F (2, 69) = .04, p = .96; post-test 2, F (2, 69) = .60, p = .55. Parametric tests were used therefore. Figure 6.4.1 below presents the mean scores of each group. The descriptive statistics follow in Table 6.4.1.

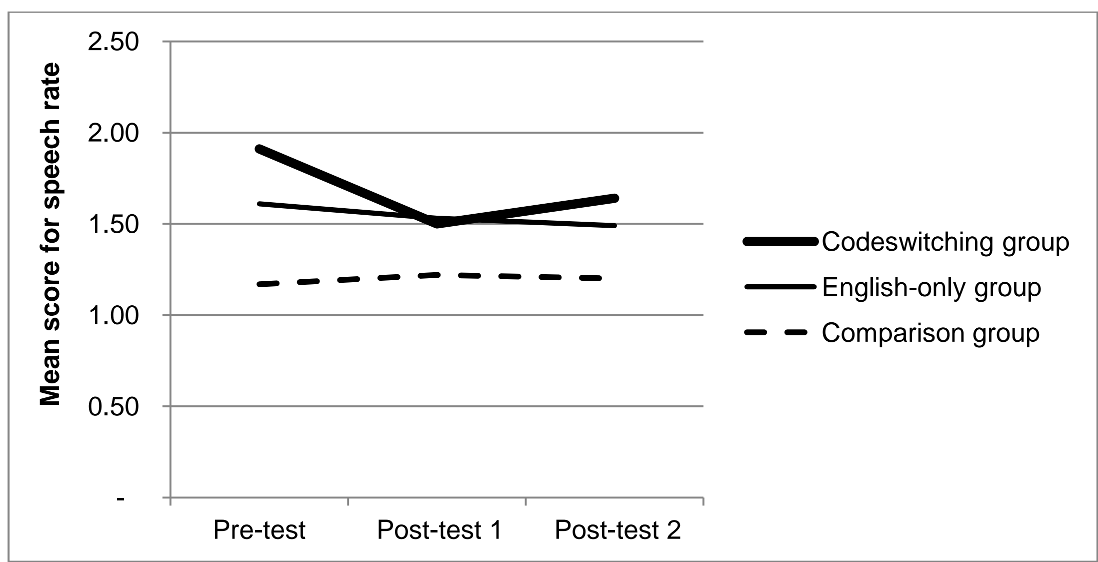


Figure 6.4.1 Line graph: Line graph: Mean performance of speech rate of the three groups across time

Table 6.4.1 Descriptive Statistics: Speech Rate

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	1.91 (.65)	2.31	3.22
	EO	34	1.61 (.52)	1.84	2.68
	COM	19	1.17 (.43)	1.49	1.81
Post-test 1 (Time 2)	**CS**	21	1.50 (.46)	2.02	2.76
	EO	34	1.53 (.48)	2.14	2.86
	COM	18	1.22 (.42)	1.64	2.12
Post-test 2 (Time 3)	**CS**	21	1.64 (.47)	1.88	2.70
	EO	35	1.49 (.48)	2.07	2.87
	COM	19	1.20 (.39)	1.44	1.76

A one-way ANOVA showed that there is a significant difference between the groups at pre-test, F (2, 73) = 9.51, $p < .001$, $r = .46$. This was close to a large effect as it explained 21% of the total variance. According to the Hochberg test, this difference lay between the CS and the COM group, $p < .001$, and between the EO and the COM group, $p < .05$, but not between the CS and the EO group, $p = .12$. Due to these differences at pre-test, two ANCOVAs were conducted for the between-group differences at time 2 and time 3, with time 1 as the covariate.

The first ANCOVA showed that the covariate, namely pre-test scores, was significantly related to post-test 1 scores of speech rate, F (1, 68) = 59.36, $p < .001$, $partial\ \eta^2 = .47$. The effect of these scores was large, as it explained 47% of the total variance. Having controlled for this effect, no significant effect of groups was found, F (2, 68) = 2.46, $p = .09$. The second ANCOVA showed similar results. Although pre-test scores were significanlty related to post-test 2 scores, F (1, 70) = 61.05, $p < .001$, $partial\ \eta^2 = .47$ (large effect explaining 47% of variance), no significant effect of groups was found, when the effect of the covariate was controlled, $p = .86$. So although at a descriptive level the graph gives the impression that the speech rate of the CS group drops, the inferential statistics confirm that there were no significant differeces between the groups.

ii. Mean Length of Runs (MLR)

The second variable in this construct is the MLR, which is the average length of runs that occurs between short pauses. The distribution graphs of the data appear in Appendix 38. According to the K-S test, these are not normally distributed: pre-test, D (72) = .16, $p < .001$; post-test 1, D (72) = .17, $p < .001$; post-test 2, D (72) = .13, $p < .01$. Nevertheless, based on the results of the Levene's test, homogeneity requirements have been met: pre-test, F (2, 69)

= 1.87, p = .16; post-test 1, F (2, 69) = 2.04, p = .14; post-test 2, F (2, 69) = 2.03, p = .14. As the assumption of normality has not been met however, the results of parametric tests were compared to those of non-parametric tests (see Appendix 40). These looked at the between-group differences at each time point. The results of the two types of tests were the same; there is a difference that approaches significance at pre-test, while there are significant differences at post-tests. As there is no equivalent of a mixed ANOVA in non-parametric tests, the mixed ANOVA was chosen to explore the data further. The mean scores for each group are presented below in Figure 6.4.2.

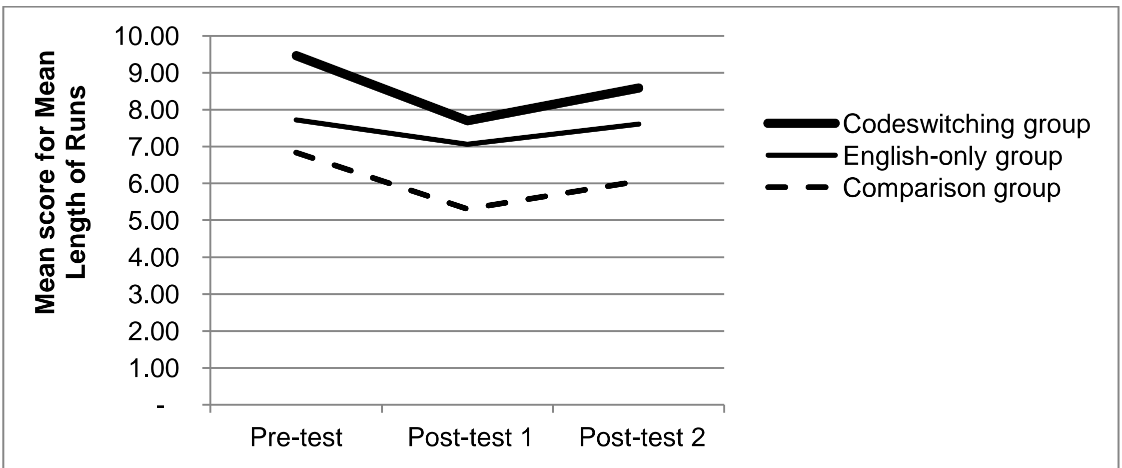


Figure 6.4.2 Line graph: Mean performance of MLR of the three groups across time

Table 6.4.2 Descriptive Statistics: Mean Length of Runs scores

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	9.46 (4.18)	14.84	18.50
	EO	34	7.72 (3.39)	18	20.57
	COM	19	6.84 (2.51)	10.46	14
Post-test 1 (Time 2)	**CS**	21	7.70 (3.15)	11.91	15.67
	EO	34	7.06 (2.68)	12.14	12.8
	COM	18	5.31 (1.83)	7	10
Post-test 2 (Time 3)	**CS**	21	8.59 (2.81)	9.18	13.92
	EO	35	7.61 (2.64)	13.26	17
	COM	19	6.06 (1.77)	6.97	10.62

A one-way ANOVA showed that the difference between the groups at pre-test is approaching significance, F (2, 73) = 3.08, p = .052, r = .28. The effect of this variable was medium as it explained 7.84% of the total variance. The Hochberg test revealed that the difference between the CS and the COM group is approaching significance, p = .055. Two ANCOVAs were conducted therefore for the between-group differences at time 2 and time 3, with time 1 as the covariate.

The first ANCOVA showed that the covariate, namely pre-test scores, was significantly related to post-test 1 scores of MLR, F (1, 68) = 18.59, $p < .001$, *partial* η^2= .21. This effect explained 21% of the total variance so it is considered to be almost a large effect. After controlling for the effect of pre-test scores, the effect of groups on post-test 1 scores were approaching significance, F (2, 68) = 2.64, p = .078, *partial* η^2 = .07. This was a medium effect as it explained 7% of the total variance. Pairwise comparisons revealed that there is a significant difference between the EO and the COM group, $p < .05$, while the difference between the CS and the COM group is approaching significance, p = .063. No significant difference was found between the CS and the EO group, p = .97.

The second ANCOVA revealed that pre-test scores are significantly related to post-test 2 scores of MLR, F (1, 70) = 30.13, $p < .001$, *partial* η^2= .30. This was an almost large effect as it explained 30% of the total variance. After controlling for the effect of pre-test scores, the effect of groups on post-test 2 scores is approaching significance, F (2, 70) = 2.85, p = .065, *partial* η^2 = .07. The effect of groups was medium, as it explained 7% of the variance. Based on pairwise comparisons, this difference lay between the CS and the COM group, $p < .05$, the EO and the COM group, $p < .05$, but not between the CS and the EO group, p = .73.

It would be interesting to note here that at first glance the line graph suggests a decrease of MLR on the part of all three groups. Although this should not be surprising for the EO and

COM group, the hypothesis predicted a fluency increase for the CS group. A reasonable explanation could be the fact that the limited intervention did not allow enough time for the CS group to go through the necessary proceduralisation processes. However, it should be stressed that the inferential statistics, which are the ones to be considered when answering the research question, suggest that, even when the initial pre-test differences were taken into account, the experimental groups did significantly better than the comparison group after the intervention.

iii. *Pause Rate*

The third variable in this construct is *pause rate,* which shows the percentage of speaking time in seconds spent in filled and unfilled pauses. The graphs in Appendix 38 show the distributions of the pause rate data as well. The K-S test indicated that the data were normally distributed at pre-test and post-test 2, but not at post-test 1: pre-test, D (72) = .08, p = .20; post-test 1, D (72) = .13, p < .01; post-test 2, D (72) = .09, p = .20. Nevertheless, the Levene's test indicated that the assumption of homogeneity of variance was not violated: pre-test, F (2, 69) = 2.13, p = .13; post-test 1, F (2, 69) = .12, p = .89; post-test 2, F (2, 69) = .19, p = .82. As the assumption of normality was not met at time 2, non-parametric tests and parametric tests explored the between-group differences of these data and their results were compared (Appendix 40). Both tests indicated a significant difference of pause rate between the groups at time 2. Since the results were the same therefore, parametric tests were used. The line graph and table below present how pause rate progressed over time.

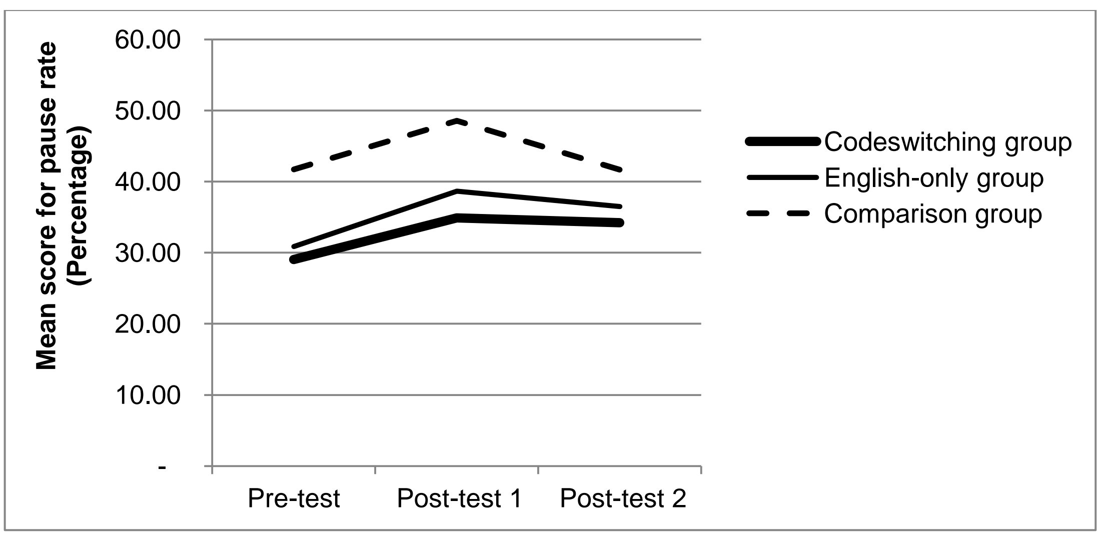


Figure 6.4.3 Line graph: Mean performance of pause rate of the three groups across time

Table 6.4.3 Descriptive Statistics: Pause Rate

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	29.06 (10.90)	50.55	52.80%
	EO	34	30.85 (15.32)	60.55	65.72%
	COM	19	41.70 (16.34)	60.20	69.85%
Post-test 1 (Time 2)	**CS**	21	34.90 (16.26)	62.86	76.39%
	EO	34	38.67 (15.15)	63.81	84.92%
	COM	18	48.56 (16.31)	52.86	69.85%
Post-test 2 (Time 3)	**CS**	21	36.20 (12.64)	43.16	60.82%
	EO	35	36.51 (13.55)	54.18	68.60%
	COM	19	41.66 (16.28)	69.55	85.21%

A one-way ANOVA showed that there is a significant difference between the three groups at pre-test, F (2, 73) = 4.93, $p < .05$, $r = .35$. The effect of this variable was medium as it explained 12.25% of the total variance. Contrasts revealed that this difference lay between the CS and the COM group, $p < .05$, as well as between the EO and the COM group, $p < .05$, but not between the CS and the EO group, $p = .95$. Therefore, two ANCOVAs were conducted for the between-group differences at the two remaining time-points.

Regarding post-test 1 scores, pre-test scores were significantly related to post-test 1 scores, F (1, 68) = 52.64, $p < .001$, *partial* $\eta^2 = .44$; a very large effect which explained 44% of the total variance. Having controlled for the effect of pre-test scores, a non-significant effect of groups on post-test 1 scores was found, F (2, 68) = .84, $p = .44$. Looking at post-test 2 scores, pre-test scores were significantly related to post-test 2 scores, F (1, 70) = 45.91, $p < .001$, *partial* $\eta^2 = .40$. This was a very large effect as well, as it explained 40% of the total variance. The effect of groups on post-test 2 scores was again non-significant, after controlling for the effect of pre-test scores, F (2, 70) = .12, $p = .88$. So although at a descriptive level the graph might show a slight increase from time 1 to time 2 and a slight decrease from time 2 to time 3, the inferential statistics indicate a non-significant difference between the three groups after the intervention.

To sum up the results so far, the only significant effect found from the temporal variables is the effect of groups on the MLR measure at times 2 and 3. This difference is located between the experimental groups and the comparison group at post-tests. The meaning of this result is discussed later on in this chapter.

iv. Repetitions

The next fluency variable is the first dysfluency marker, namely repetitions. Appendix 39 contains the distribution graphs of each time point. According to the K-S test, these are not normally distributed: pre-test, D (72) = .21, $p < .001$; post-test 1, D (72) = .21, $p < .001$; post-test 2, D (72) = .17, $p < .001$. The Levene's test indicated that some datasets were not homogenous either: pre-test, F (2, 69) = 2.89, $p = .062$; post-test 1, F (2, 69) = 5.09, $p < .01$; post-test 2, F (2, 69) = 1.66, $p = .19$. The results of non-parametric tests were compared to those of parametric tests therefore, in order to determine the extent to which these differ

(Appendix 40). The results of both tests showed that there were no between-group differences at any time point. Since these were the same, parametric tests were used. The line graph below shows how the amount of repetitions increased or decreased throughout the intervention, while Table 6.4.4 gives the exact figures.

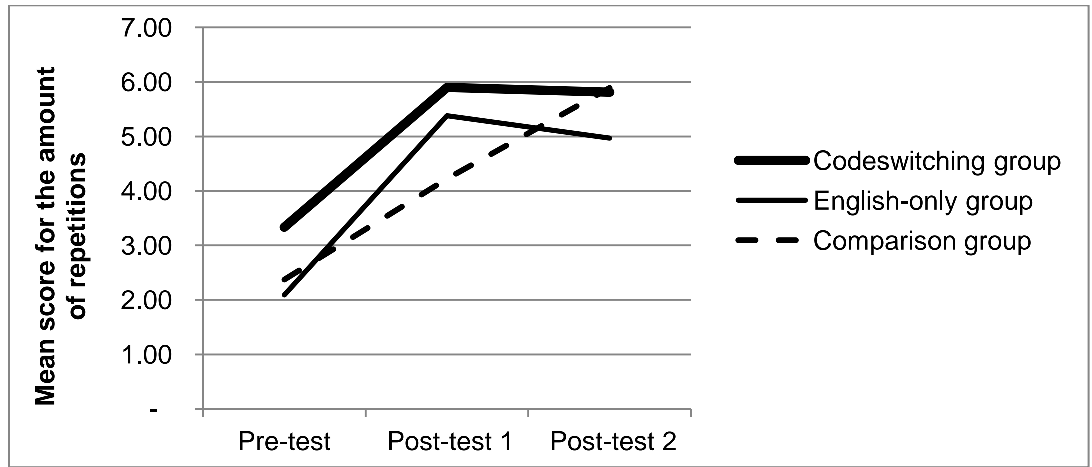


Figure 6.4.4 Line graph: Mean performance of the amount of repetitions of the three groups across time

Table 6.4.4 Descriptive Statistics: Repetitions

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	3.33 (2.80)	10	10
	EO	34	2.09 (1.96)	10	10
	COM	19	2.37 (2.36)	10	10
Post-test 1 (Time 2)	**CS**	21	5.90 (6.69)	28	29
	EO	34	5.38 (6.22)	14	14
	COM	18	6.22 (2.24)	9	9
Post-test 2 (Time 3)	**CS**	21	5.81 (6.13)	18	18
	EO	35	4.97 (4.69)	18	18
	COM	19	5.89 (4.77)	18	18

A one-way ANOVA showed that there is no significant difference between the groups at pre-test, F (2, 73) = 1.91, p = .15. A mixed ANOVA explored the effects of the model further.

Looking at the within-group differences first, the Mauchly's test showed that the assumption of sphericity was met, χ^2 (2) = .39, p = .82. A significant main effect of time was found, F (2, 138) = 19.11, p < .001, *partial* η^2 = .22. This main effect was large as it explained 22% of the total variance. This suggested that the time of the test significantly affected the amount of repetitions. Contrasts revealed that the difference lay only between time 1 and time 2, F (1, 69) = 25.97, p < .001, *partial* η^2 = .27. This was a large effect, as it explained 27% of the variance. Pairwise comparisons suggested a significant difference between time 1 and time 3 as well, p < .001. As for the between-group effect, no significant effect was found for the groups variable, F (2, 69) = .45, p = .64. This suggests that the amount of repetitions did not differ according to groups. The effect of the interaction between the time and group variable was also found to be non-significant, F (4, 138) = .74, p = .56. This means that the change of the amount of repetitions across time did not differ according to groups. So although the line graph may suggest that at a descriptive level the experimental groups reduced repetitions from time 2 to time 3, while the comparison group kept increasing the amount of repetitions, the inferential statistics confirm that there were no significant differences between the progress of the three groups.

v. <u>Reformulations</u>

The second dysfluency marker is reformulations.The graphs showing the distribution of the data at the three time points are included in Appendix 39. The K-S test indicated that these were not normally distributed: pre-test, D (72) = .21, p < .001; post-test 1, D (72) = .21, p < .001; post-test 2, D (72) = .17, p < .001. The Levene's test indicated however that they were homogenous: pre-test, F (2, 69) = 2.68, p = .08; post-test 1, F (2, 69) = .97, p = .39; post-test 2, F (2, 69) = .62, p = .54. As the assumption of normality had not been met, the results of non-parametric tests to those of parametric tests were compared (Appendix 40). The results

of the two types of tests agreed that there were no significant between-group differences at any of the time points. Due to this agreement, parametric tests were used for this variable. Figure 6.4.5 below shows how the amount of reformulations changed over time and across groups, while Table 6.4.5 gives the descriptive statistics of this variable.

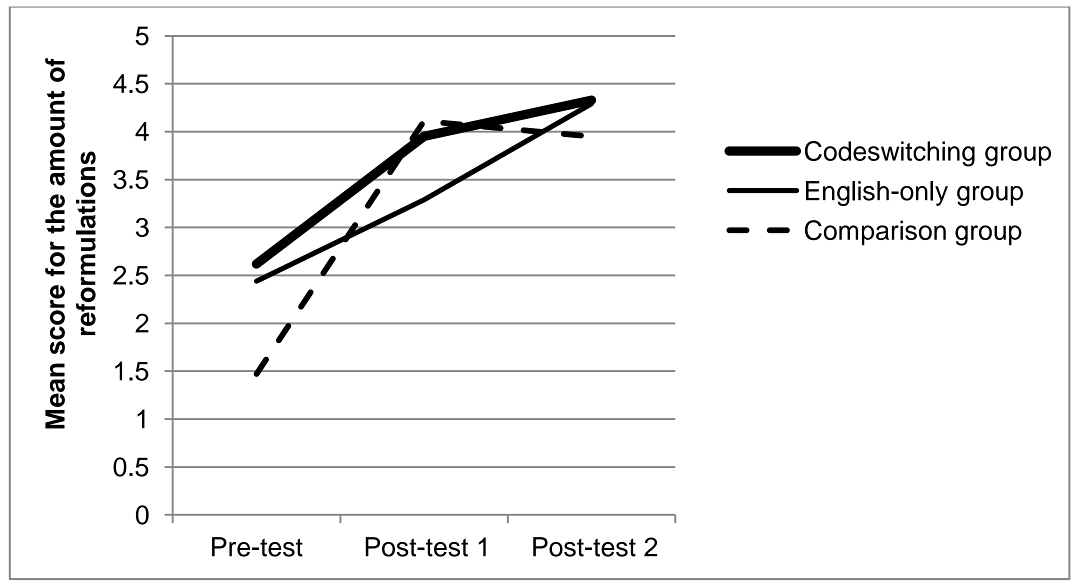


Figure 6.4.5 Line graph: Mean performance of the amount of reformulations of the three groups across time.

Table 6.4.5 Descriptive Statistics: Reformulations

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	2.62 (2.80)	10	10
	EO	34	2.44 (2.06)	8	8
	COM	19	1.47 (1.64)	5	5
Post-test 1 (Time 2)	**CS**	21	3.95 (3.60)	11	11
	EO	34	3.29 (2.93)	14	14
	COM	18	4.11 (2.47)	7	8
Post-test 2 (Time 3)	**CS**	21	6.33 (3.97)	14	14
	EO	35	6.29 (2.92)	10	10
	COM	19	3.95 (3.34)	13	13

A one-way ANOVA showed that there was no significant difference between the three groups at pre-test, F (2, 73) = 1.59, p = .21. A mixed ANOVA was conducted therefore to explore the effect in the model. In terms of sphericity, the Mauchly's test indicated that this assumption had been met, χ^2 (2) = .42, p = .81. A main effect of time was found to be significant, F (2, 138) = 16.37, $p < .001$, *partial* η^2 = .17. This was a big effect, as it explained 17% of the total variance. This suggests that the time of the test significantly affected the amount of reformulations. Contrasts revealed that the difference lay only between time 1 and time 2, F (1, 69) = 17.47, $p < .001$, *partial* η^2 = .20; a large effect that explained 20% of the total variance. Pairwise comparisons revealed a significant difference between time 1 and time 3 as well, $p < .001$.

As for the between-group effect, a non-significant main effect was found, F (2, 69) = .25, p = .78. This suggests that the amount of reformulations did not differ according to groups. Similarly, a non-significant interaction effect between the time of the test and groups was found, F (4, 138) = 1.11, p = .35. This indicates that the difference of the amount of reformulations across time did not differ according to groups.

It is worth noting that the line graph may initially suggest that the experimental groups showed a continuous increase from time 2 to time 3, while the comparison group reduced its use of reformulations. However, as with the repetitions variable, the inferential statistics indicated that there was no significant difference between the progress of the three groups.

vi. <u>*Incomplete words*</u>

The final dysfluency marker is incomplete words. The distribution graphs for this variable are included in Appendix 39. The K-S test indicated that these were not normally distributed: pre-test, D (72) = .25, p = .001; post-test 1, D (72) = .17, p < .001; post-test 2, D (72) = .17, p < .001. As for homogeneity of variance, the Levene's test indicated that this had not been met at all time points: pre-test, F (2, 69) = .96, p = .39; post-test 1, F (2, 69) = 5.53, p < .01; post-test 2, F (2, 69) = 3.51, p < .05. Non-parametric and parametric tests were conducted therefore, in order to compare the results (Appendix 40). No significant between-group differences were found at any time point in both types of tests. Since the results were the same therefore, parametric tests were used in order to perform mixed design tests. Figure 6.4.6 below shows how this amount changed over time and across groups. The descriptive statistics are presented in Table 6.4.6.

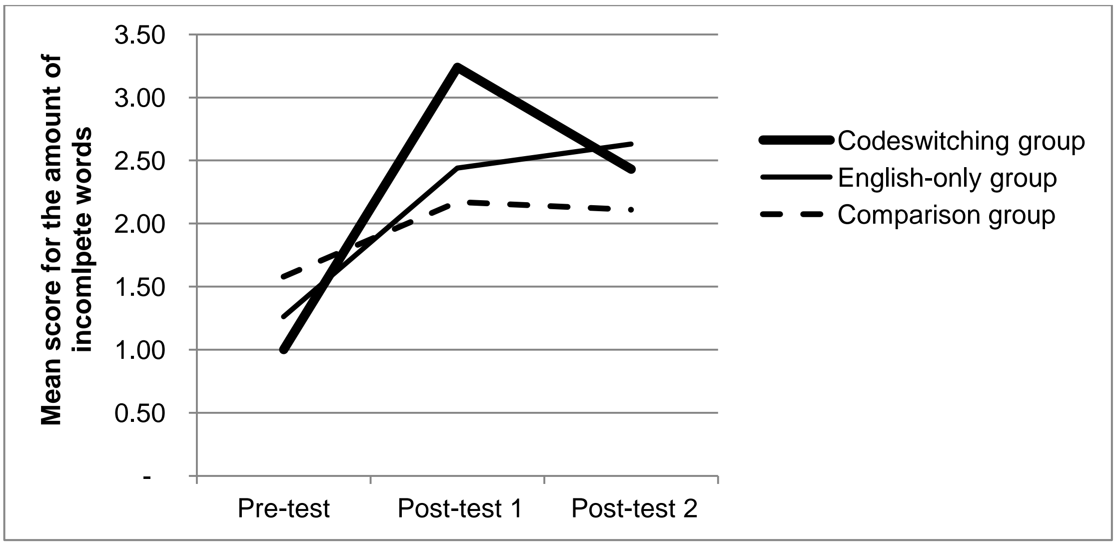


Figure 6.4.6 Line graph: Mean performance of the amount of incomplete words of the three groups across time.

Table 6.4.6 Descriptive Statistics: Incomplete words

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	1.00 (1.97)	8	8
	EO	34	1.26 (1.33)	4	4
	COM	19	1.58 (1.71)	4	4
Post-test 1 (Time 2)	**CS**	21	3.24 (4.15)	13	13
	EO	34	2.44 (2.29)	7	7
	COM	18	2.17 (1.42)	4	4
Post-test 2 (Time 3)	**CS**	21	2.43 (2.67)	9	9
	EO	35	2.63 (2.03)	7	7
	COM	19	2.11 (1.88)	7	7

As with the other dysfluency markers, no significant difference was found between the three groups at pre-test, F (2, 73) = .63, p = .54. A mixed ANOVA therefore was conducted for this variable as well.

For the within-group main effect, the Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (2) = 6.66, p < .05. Looking at the Greenhouse-Geisser test therefore, a significant effect of time was found, F (1.83, 3.66) = 10.29, p < .001, *partial* η^2 = .13. This was a small effect, as it only accounted for 13% of the total variance. Contrasts revealed that the difference lay between time 1 and time 2 only, F (1, 69) = 17.90, p < .001, *partial* η^2 = .21; a large effect which accounted for 21% of the variance. According to pairwise comparisons, the difference between time 1 and time 3 was significant as well, p < .001. As for the between-group effect, a non-significant effect of groups was found, F (2, 69) = .30, p = .74. The amount of incomplete words therefore did not differ according to groups. No significant interaction effect between time and groups was found on incomplete words, F

(4, 138) = 1.26, p = .29. This indicates that the change of the amount of incomplete words across time did not differ according to groups.

As with the previous dysfluency markers, the line graph shows a slight decline on the part of the CS and the COM group from time 2 to time 3 and a slight increase on the part of the EO group. However, the inferential statistics confirmed that these differences were non-significant and therefore will not be considered when answering the research question.

The results of the three dysfluency markers are quite similar. The inferential statistics suggest that, while there is a significant within-group difference between pre-test and post-tests, being in a group does not affect the amount of dysfluency markers. Also, the non-significant interaction effects suggest that this change of dysfluency markers across time does not differ according to groups. Due to this lack of significant effects therefore, no claims can be made in response to RQ4 based on dysfluency markers.

6.4.2 Content variables

The second construct of variables measures quality of content. Three variables were selected for this category: *verb ratio, L1 syllables* and *idea units.*

i. <u>*Verb Ratio*</u>

The verb ratio looks at the diversity of verb phrase use. It was calculated by dividing the number of different types of verbs by the total number of verbs used, and then multiplying this figure by 100. The higher the figure is, the greater lexical variety is being used. The distribution graphs are included in appendix 41. These look normally distributed and this was confirmed by the K-S test as well: pre-test, D (72) = .09, p = .09; post-test 1, D (72) = .08, p = .20; post-test 2, D (72) = .05, $p < .05$. According to the Levene's test, these were also

homogenous: pre-test, F (2, 69) = 2.34, p = .10; post-test 1, F (2, 69) = .76, p = .47; post-test 2, F (2, 69) = 2.79, p = .07. Since both assumptions have been met, parametric tests were used.

Figure 6.4.7 below shows the mean scores of each group, while Table 6.4.7 presents the descriptive statistics.

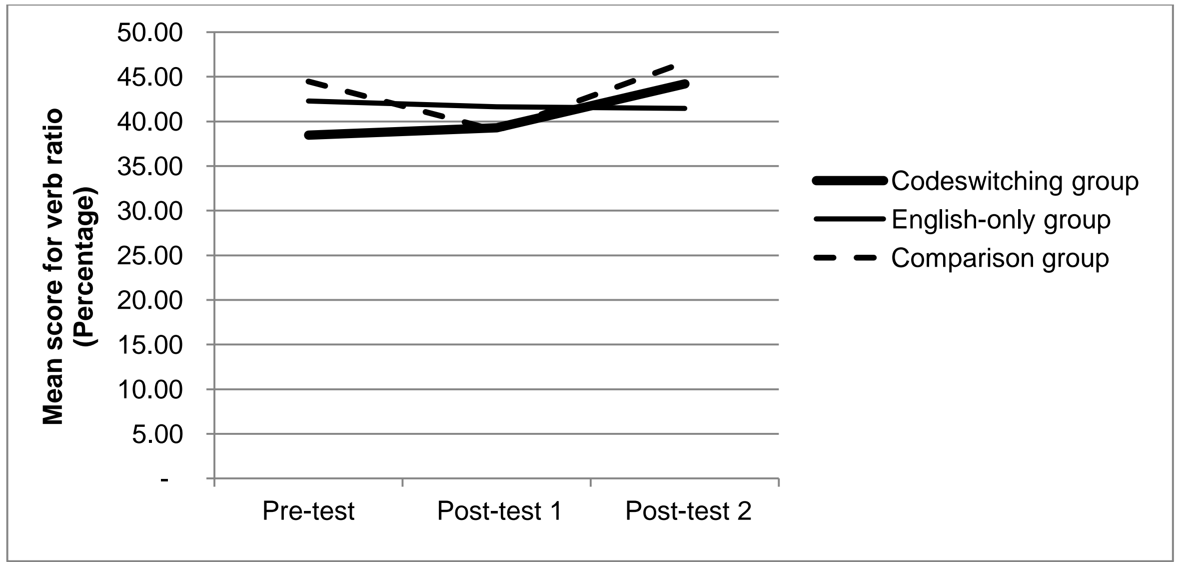


Figure 6.4.7 Line graph: Mean performance of verb ratio of the three groups across time.

Table 6.4.7 Descriptive Statistics: Verb Ratio

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	38.48 (7.66)	29.54	56.54%
	EO	34	42.28 (11.72)	56.06	83.33%
	COM	19	46.49 (13.90)	53.33	70%
Post-test 1 (Time 2)	**CS**	21	39.28 (9.70)	39	60%
	EO	34	41.65 (8.71)	41	69%
	COM	18	39 (10.35)	36	58%
Post-test 2 (Time 3)	**CS**	21	46.22 (9.24)	35.72	62.50%
	EO	35	41.45 (6.42)	26.59	56.41%
	COM	19	46.64 (9.22)	29.76	58.33%

A one-way ANOVA showed that the difference between the three groups at pre-test was non-significant, F (2, 73) = 1.46, p = .24. A mixed ANOVA was conducted therefore. As the assumption of sphericity had been violated, χ^2 (2) = 9.11, $p < .05$, the Greenhouse-Geisser test revealed that the effect of time was approaching significance, F (1.78, 122.62) = 3.10, p = .055, *partial* η^2 = .04. This was a small to medium effect as it explained 4% of the total variance. Contrasts revealed that this difference lay between time 2 and time 3, F (1, 69) = 9.56, $p < .01$, *partial* η^2 = .12, but not between time 1 and time 2, F (1, 69) = 1.25, p = .27. Pairwise comparisons revealed that there is no significant difference between time 1 and time 3 either, p = .24. As for the effect of groups, no significant effect was found, F (2, 69) = 1.11, p = .34. This means that verb ratio did not differ according to groups. No significant interaction effect was found either, F (4, 138) = 1.68, p = .16. The change of verb ratio across time therefore did not differ according to groups.

ii. *L1 syllables*

The second variable in this construct, namely *L1 syllables,* represents the total number of L1 syllables used by a student. This variable was included in the analysis for two reasons; first, to examine whether allowing codeswitching leads to excessive L1 use; and second, to ensure that any potential fluency gains are representative of L2 capacity, in the sense that they are based on L2 speech, rather than a mixture of two languages.

The distribution graphs are presented in Appendix 41. According to the K-S test, these were significantly not normally distributed: pre-test, D (72) = .24, $p < .001$; post-test 1, D (72) = .23, $p < .001$; post-test 2, D (72) = .22, $p < .001$. The Levene's test indicated that not all sets of data met the assumption of homogeneity of variance: pre-test, F (2, 69) = 1.23, p = .29; post-test 1, F (2, 69) = 3.54, $p < .05$; post-test 2, F (2, 69) = 5.89, $p < .01$. Therefore, the

comparison of the results of non-parametric tests to those of parametric tests was conducted for this variable as well (Appendix 43). Both types of tests indicated that there were no significant between-group differences at any of the time points. Since these were the same therefore, parametric tests were used in order to explore mixed design effects. Figure 6.4.8 below shows the mean scores for each group, while the descriptive statistics follow.

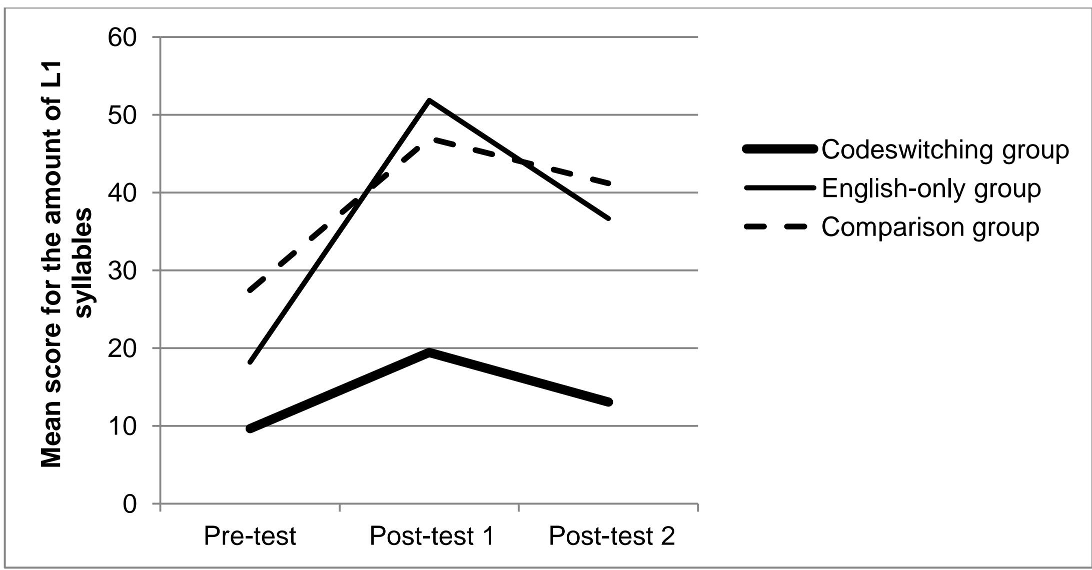


Figure 6.4.8 Line graph: Mean performance of the amount of L1 syllables of the three groups across time.

Table 6.4.8 Descriptive Statistics: L1 Syllables

	Group	N	Mean (SD)	Range	Max
Pre-test	**CS**	21	9.62 (16.32)	71	71
	EO	35	18.17 (24.15)	90	90
	COM	19	27.42 (36.91)	144	144
Post-test 1	**CS**	21	19.43 (15.76)	55	57
	EO	34	51.82 (64.17)	269	272
	COM	18	46.94 (55.18)	238	241
Post-test 2	**CS**	21	13.05 (11.80)	37	37
	EO	35	36.66 (46.49)	180	180
	COM	19	41.16 (59.08)	232	232

A one-way ANOVA showed that there is no significant difference between the groups at pre-test, F (2, 72) = 2.61, p = .08. A mixed ANOVA was conducted therefore, in order to further explore this variable between groups and across time.

As far as the within-group effect is concerned, the Mauchly's test showed that the assumption of sphericity had been violated, χ^2 (2) = 36.51, $p < .001$. Looking at the Greenhouse-Geisser test therefore, a significant effect of time on the amount of L1 syllables was revealed, F (1.43, 2.87) = 16.90, $p < .001$, *partial* $\eta^2 = .19$. This was a medium to large effect as it accounted for 19% of the total variance. Contrasts revealed that this difference lay between time 1 and time 2, F (1, 70) = 20.77, $p < .001$, *partial* $\eta^2 = .23$ (large effect accounting for 23% of variance), as well as between time 2 and time 3, F (1, 70) = 13.56, $p < .001$, *partial* $\eta^2 = .16$ (medium to large effect accounting for 16% of the variance). According to pairwise comparisons, the difference between time 1 and time 3 was also significant, $p < .05$. As for the between-group differences, a non-significant effect was found, F (2, 70) = 2.65, p = .08. This suggests that, when the time variable is disregarded, L1 syllables do not differ according to groups. No significant interaction effect was found either, F (4, 140) = 1.85, p = .12. This indicates that the change of L1 syllables across time did not differ according to groups.

It is worth mentioning here that the fact that the CS group did not use significantly more L1 syllables than the other two groups supports the argument that the incorporation of codeswitching in L2 pedagogy does not necessarily lead to excessive and uncontrollable L1 use. This argument will be further discussed in the Discussion chapter.

iii. Idea units

The final measure of quality of content is idea units. This was calculated by counting the number of propositions made in the presentation part of the oral production tests. Graphs with the distributions of the amount of idea units at the three time points are included in Appendix 41. Although these look satisfactorily normal, the K-S test suggested that they were not: pre-test, D (72) = .13, $p < .01$; post-test 1, D (72) = .20, $p < .001$; post-test 2, D (72) = .16, $p < .001$. The Levene's test indicated that the data were homogenous at times 1 and 3, but not at time 2: pre-test, F (2, 69) = .06, $p = .94$; post-test 1, F (2, 69) = 6.51, $p < .05$; post-test 2, F (2, 69) = 1.19, $p = .31$. Comparisons of the results of non-parametric to parametric tests were made therefore (Appendix 43). As these agreed to a great extent, parametric tests were used for idea units. The figure and table below present the descriptive statistics of idea units.

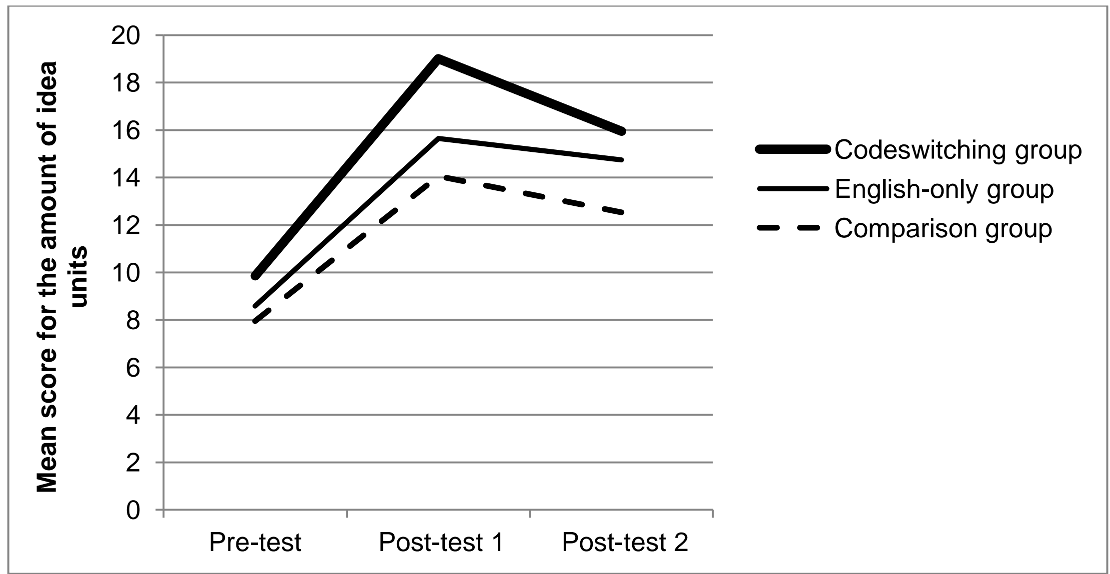


Figure 6.4.9 Line graph: Mean performance of the amount of idea units of the three groups across time.

Table 6.4.9 Descriptive Statistics: Idea units

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	9.86 (3.17)	10	16
	EO	34	8.59 (2.94)	12	16
	COM	19	7.95 (2.59)	7	12
Post-test 1 (Time 2)	**CS**	21	19 (8.30)	30	41
	EO	34	15.65 (4.90)	23	31
	COM	18	14.06 (3.02)	12	21
Post-test 2 (Time 3)	**CS**	21	15.95 (7.57)	35	40
	EO	35	14.74 (6.32)	19	25
	COM	19	12.53 (4.09)	19	25

A one-way ANOVA showed that there are no significant pre-test differences between the three groups in the amount of idea units, F (2, 73) = 2.27, p = .11. A mixed ANOVA was conducted therefore.

For the within-group effect, the Mauchly's test indicated that the assumption of sphericity had been violated, χ^2 (2) = 6.47, $p < .05$. According to the Greenhouse-Geisser test, there was a significant effect of time on the amount of idea units, F (1.83, 3.67) = 94.02, $p < .001$, *partial* η^2 = .58. According to Cohen (1988, 1992), this was a very large effect, as it explained 50% of the variance. Contrasts revealed that the difference lay between time 1 and time 2, F (1, 69) = 161.95, $p < .001$, *partial* η^2 = .70 (large effect explaining 70% of variance), as well as between time 2 and time 3, F (1, 69) = 15.18, $p < .001$, *partial* η^2 = .18 (medium to large effect explaining 18% of the variance). Pairwise comparisons revealed that the difference between time 1 and time 3 was also significant, $p < .001$.

As for the between-group effect, a significant effect of groups was found, F (2, 69) = 3.39, $p < .05$, *partial* η^2 = .09. This was a medium effect that explained 9% of the total variance.

Pairwise comparisons revealed that this difference laid between the CS and the COM group only, $p < .05$. A non-significant interaction effect was found, F (4, 138) = 1.63, $p = .17$. This means that the change of idea units across time did not differ according to groups.

Looking at the results of the content variables collectively, they were all significantly affected by the time of the test. The verb ratio significantly increased from time 2 to time 3, while the L1 syllables and the idea units showed a significant increase from time 1 to time 2 and significant decrease from time 2 to time 3. A significant effect of groups was also found on the amount of idea units. The CS group expressed significantly more idea units overall than the COM group. No interaction effect was found on any of the three variables. These results are interpreted at the end of the analysis for this research question.

6.4.3 Accuracy variables

The final category of variables aims to measure accuracy. It consists of two variables: *percentage of error-free clauses* and *percentage of correct verb use.*

i. <u>*Percentage of error-free clauses*</u>

Looking at the percentage of error-free clauses first, the distribution graphs of the data are shown in Appendix 42. Although these look satisfactorily normal, the K-S test indicated that they were not normally distributed at time 2 and 3: pre-test, D (72) = .08, $p = .20$; post-test 1, D (72) = .11, $p < .05$; post-test 2, D (72) = .12, $p < .01$. As for homogeneity of variance, the Levene's test showed that this assumption had been met: pre-test, F (2, 69) = .26, $p = .77$; post-test 1, F (2, 69) = .01, $p = .99$; post-test 2, F (2, 69) = .18, $p = .84$. Therefore, non-parametric tests were compared to parametric tests for time 2 and 3 only; the time points with non-normal data (Appendix 43). The results agree on a non-significant between-group

difference at time 2, and a significant difference at time 3. Due to this agreement, parametric tests can be used for further exploration of the data. The line graph and table below present the results for the percentage of error-free clauses first.

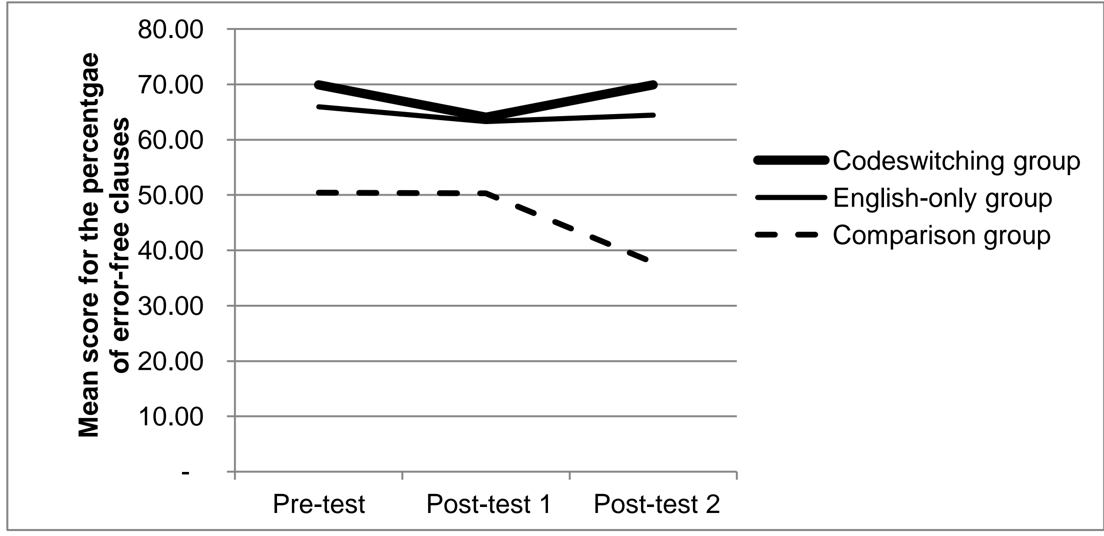


Figure 6.4.10 Line graph: Mean performance of the percentage of error-free clauses of the three groups across time.

Table 6.4.10 Descriptive Statistics: Percentage of error-free clauses

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	69.90 (22.98)	100	100%
	EO	34	65.94 (24.02)	100	100%
	COM	19	50.42 (28.17)	100	100%
Post-test 1 (Time 2)	**CS**	21	64.04 (25.24)	82.35	100%
	EO	34	63.28 (28.44)	100	100%
	COM	18	50.33 (27.09)	100	100%
Post-test 2 (Time 3)	**CS**	21	69.92 (22.65)	83.33	100%
	EO	35	66.43 (26.91)	100	100%
	COM	19	37.81 (26.33)	83.33	83.33%

A one-way ANOVA at pre-test revealed a significant difference between the groups in terms of the percentage of error-free clauses, F (2, 73) = 3.48, $p < .05$, $r = .30$. This effect was

medium as it accounted for 9% of the variance. Contrasts revealed that this difference lay between the CS and the COM group only. Due to this difference, two ANCOVAs explored the between-group differences at time 2 and time 3.

According to the ANCOVAs, the covariate was significantly related to post-test 1 scores, F (1, 68) = 43.69, $p < .001$, *partial* $\eta^2 = .39$; a large effect that explained 39% of the variance. Having controlled for the effect of pre-test scores, no significant effect of groups was found, $p = .64$. The covariate was also significantly related to post-test 2 scores, F (1, 70) = 21.11, $p < .001$, *partial* $\eta^2 = .23$. Again, according to Cohen (1988), this was a large effect as it accounted for 23% of the total variance. After controlling for the effect of pre-test scores, the group variable was also significantly related to post-test 2 scores, F (2, 70) = 6.49, $p < .01$, *partial* $\eta^2 = .16$; this was a medium to large effect which explained 16% of the variance. Pairwise comparisons revealed that the difference lay between the CS and the COM group, $p < .01$, the EO and the COM group, $p < .01$, but not between the CS and the EO group, $p = .57$.

ii. Percentage of correct verb use

Moving on to the second variable of accuracy, the distribution graphs for the percentage of correct verb use are shown in Appendix 42. According to the K-S test, the assumption of normality was not met: pre-test, D (72) = .21, $p < .001$; post-test 1, D (72) = .11, $p < .05$; post-test 2, D (72) = .14, $p = .001$. The Levene's test showed that the assumption of homogeneity of variance was met: pre-test, F (2, 69) = 2.56, $p = .08$; post-test 1, F (2, 69) = .28, $p = .75$; post-test 2, F (2, 69) = 1.13, $p = .33$. As the normality requirements were not met, the results of non-parametric to parametric tests on the data were compared (Appendix 43). Again the results of these two tests were the same. Parametric tests therefore were used

for the analysis of this variable. The line graph below shows the progress of the three groups in terms of correct use of verbs. Table 6.4.11 provides the exact figures.

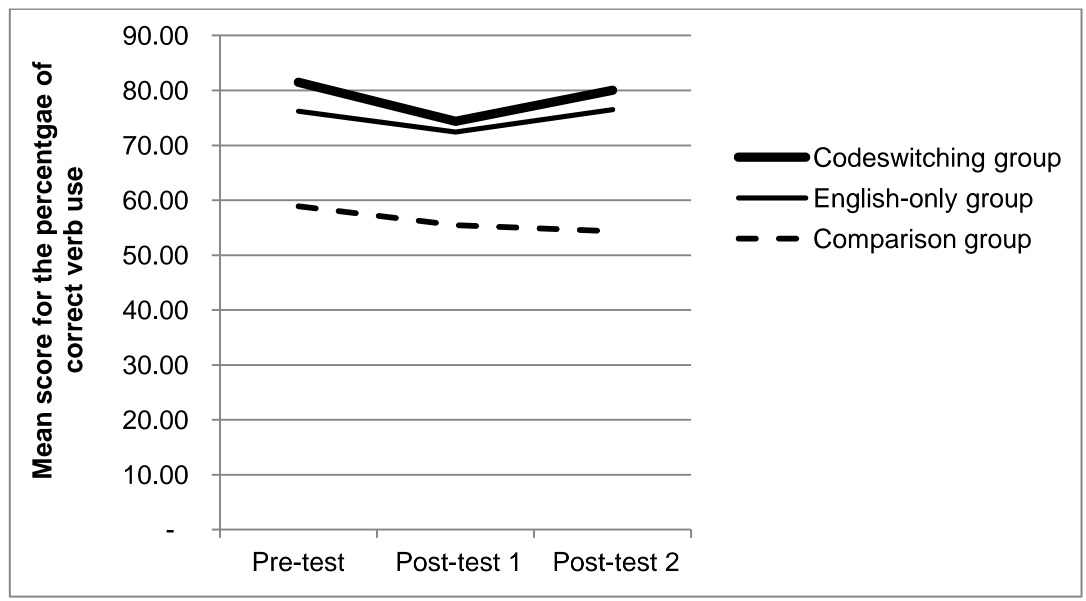


Figure 6.4.11 Line graph: Mean performance of the percentage of correct verbs used by the three groups across time.

Table 6.4.11 Descriptive Statistics: Percentage of correct verb use

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**CS**	21	81.51 (23.63)	100	100%
	EO	34	76.24 (23.40)	100	100%
	COM	19	58.78 (30.72)	100	100%
Post-test 1 (Time 2)	**CS**	21	76.38 (19.52)	66.67	100%
	EO	34	72.42 (26.55)	100	100%
	COM	18	55.46 (26.04)	100	100%
Post-test 2 (Time 3)	**CS**	21	80.04 (19.48)	80	100%
	EO	35	76.50 (21.88)	100	100%
	COM	19	56.34 (27.07)	87.10	87.10%

A significant difference was found between the three groups at pre-test, F (2, 73) = 4.39, p < .05, r = .33. Contrasts revealed that this difference lay between the CS and the COM group, p

$< .05$, while the difference between the EO and the COM group is approaching significance, $p = .057$. No significant difference was found between the CS and the EO group, $p = .84$. Due to these differences, two ANCOVAs were conducted for the between-group differences at Time 2 and Time 3.

The first ANCOVA revealed that the covariate was significantly related to post-test 1 scores, $F\ (1, 68) = 52.59$, $p < .001$, $partial\ \eta^2 = .44$. This was a large effect that accounted for 44% of the total variance. Having controlled for the effect of pre-test scores, no significant effect of groups was found, $F\ (2, 68) = 1.84$, $p = .17$. The second ANCOVA revealed that the covariate was also significantly related to post-test 2 scores, $F\ (1, 70) = 34.79$, $p < .001$, $partial\ \eta^2 = .33$; a large effect that explained 33% of the variance. After controlling for the effect of pre-test scores, the group variable was significantly related to post-test 2 scores, $F\ (2, 70) = 5.33$, $p < .01$, $partial\ \eta^2 = .13$. This was a medium effect, as it explained 13% of the total variance. Pairwise comparisons revealed that the difference lay between the CS and the COM group, $p < .05$, and the EO and the COM group, $p < .05$, but not between the CS and the EO group, $p = .80$.

The two accuracy variables yield similar results. While no significant difference is found immediately after the intervention, the delayed post-test indicates that there were no detrimental effects on the accuracy of the two experimental groups. These results are discussed in the following section.

6.4.4. Response to RQ4

Taking all the above results into consideration, this section attempts to give a response to the fourth research question, which first examines whether TR+ favours fluency development, in relation to no task repetition; and second, whether task repetition with codeswitching *plus*

leads to more fluency development than task repetition with English-only *plus*. This inquiry assumes no detrimental effects on accuracy. The results of the fluency variables are considered here in relation to the results of the content variables and those of the accuracy variables. The four tables below show the results of all these variables collectively.

Table 6.4.12 Results of temporal variables

FLUENCY MEASURES (Temporal variables)		
Speech Rate	2 ANCOVAs	
	Pre-test	*sig**** CS & COM*** EO & COM*
	Post-test 1	*ns*
	Post-test 2	*ns*
Mean Length of Runs	2 ANCOVAs	
	Pre-test	*sig* appr. CS & COM appr.
	Post-test 1	*sig* appr. CS & COM appr. EO & COM*
	Post-test 2	*sig* appr. CS & COM* EO & COM*
Pause rate	2 ANCOVAs	
	Pre-test	*sig** CS & COM* EO & COM*
	Post-test 1	*ns*
	Post-test 2	*ns*

Table 6.4.13 Results of dysfluency markers

FLUENCY MEASURES (Dysfluency markers)		
Repetitions	Mixed ANOVA	
	Time	*sig**** 1 & 2*** 1 & 3***
	Groups	*ns*
	Time*Groups	*ns*
Reformulations	Mixed ANOVA	
	Time	*sig**** 1 & 2*** 1 & 3***
	Groups	*ns*
	Time*Groups	*ns*
Incomplete words	Mixed ANOVA	
	Time	*sig**** 1 & 2*** 1 & 3***
	Groups	*ns*
	Time*Groups	*ns*

sig = significant; *ns* = non-significant; $p < .05$*; $p < .01$**; $p < .001$***; appr. = approaching significance

Table 6.4.14 Results of content variables

CONTENT VARIABLES		
Verb Ratio	Mixed ANOVA	
	Time	*sig*$^{appr.}$ 2 & 3**
	Groups	*ns*
	Time*Groups	*ns*
L1 syllables	Mixed ANOVA	
	Time	*sig**** 1&2*** 2&3*** 1&3*
	Groups	*ns*
	Time*Groups	*ns*
Idea units	Mixed ANOVA	
	Time	*sig**** 1 & 2*** 2 & 3*** 1 & 3***
	Groups	*sig** CS & COM*
	Time*Groups	*ns*

sig = significant; *ns* = non-significant; $p < .05$*; $p < .01$**; $p < .001$***; appr. = approaching significance

Table 6.4.15 Results of accuracy variables

ACCURACY VARIABLES		
Percentage of error-free clauses	2 ANCOVAs	
	Pre-test	*sig** CS & COM
	Post-test 1	*ns*
	Post-test 2	*sig*** CS & COM** EO & COM**
Percentage of correct verb use	2 ANCOVAs	
	Pre-test	*sig** CS & COM* EO & COM[appr.]
	Post-test 1	*ns*
	Post-test 2	*sig*** CS & COM** EO & COM**

sig = significant; *ns* = non-significant; $p < .05$*; $p < .01$**; $p < .001$***; appr. = approaching significance

i. <u>Task repetition plus vs. no task repetition</u>

In terms of the six fluency variables, the only significant results found concerned the MLR measure at time 2 and 3 and the significant effect of time on the dysfluency markers. As discussed earlier in this chapter, the latter effect cannot contribute to providing an answer to RQ4 because of the lack of significant group and interaction effects. Therefore, the discussion will focus on the significant group effect of MLR, which was found at post-tests. The fact that the experimental groups ended up with a higher level of MLR than the comparison group, even when their pre-test differences were accounted for, is indicative of an effect. Based on the MLR only therefore, which is considered by some researchers (e.g. Towell *et al.,* 1996) to be the best indicator of fluency, it could be argued that repeating tasks and receiving feedback may lead to greater fluency, than not repeating the task at all.

Considering these results in relation to the content variables, a significant effect of time was found on all content variables, while a significant effect of groups was found between the CS and the COM group in terms of idea units. However, since this difference of idea units was not found to be significant between the EO and the COM group as well, a claim favouring TR+ over no task repetition in terms of content cannot be supported. Moreover, the absence of significant interaction effects on all content variables does not allow any claims to be made in response to RQ4, as it suggests that this significant difference of content variables across time did not differ according to groups. Attention should be drawn however to the results of the L1 syllables variable. The fact that the CS group did not have significantly more L1 syllables than the other two groups on either test suggests that L1 incorporation does not lead to excessive L1 use.

As for accuracy, both measures revealed a significant difference between the experimental and the comparison groups at post-test 2, even when the initial pre-test differences were accounted for. It could be argued therefore, that TR+ did not have a detrimental effect on accuracy.

In conclusion therefore, TR+ resulted in better fluency (based on MLR only) than no task repetition, without undermining accuracy.

ii. <u>Task repetition with CS plus vs. Task repetition with EO plus</u>

In response to the second inquiry of this research question, which aims to examine the effects of the two pedagogical packages on fluency, no results can be reported. The reason is because no significant differences between the two experimental groups were found on any of the fluency, content or accuracy measures, as well as no significant interaction effects.

The fact that the EO group did not end up being significantly more accurate than the CS group, despite the accuracy feedback, should be taken into consideration. Based on the results of the qualitative analysis of student interactions in the previous chapter however, this should not be a surprising result. This analysis suggested that it was hard for accuracy feedback to be immediately incorporated by the students. Perhaps more practice and more feedback was needed to improve accuracy. This is further discussed in the Discussion chapter.

6.5. RQ5: Do higher proficiency students benefit more from task repetition *plus* feedback than do lower proficiency students?

The final research question aims to examine whether proficiency has a role to play when it comes to the potential benefits of TR+. In order to examine this inquiry, the performance of high proficiency students *(HP group)* and low proficiency students *(LP group),* which were allocated in these groups in the way it was described in the Methodology chapter, was compared in terms of all the variable constructs. All three variables constructs were considered in order to track any potential trade-off effects, in case of a fluency change.

6.5.1. Fluency variables

i. Speech rate

The first variable in the fluency variables construct is speech rate. Graphs showing the distribution of the data at each time point are included in Appendix 44. According to the K-S test, the data were normally distributed: pre-test, D (48) = .07, p = .20; post-test 1, D (48) = .06, p = .20; post-test 2, D (48) = .09, p = .20. The Levene's test indicated that the assumption of homogeneity was met as well: pre-test, F (1, 46) = .89, p = .35; post-test 1, F (1, 46) = .19,

p = .67; post-test 2, F (1, 46) = .64, p = .43. Since both assumptions have been met, parametric tests were used. The line graph and table below show the progress of the two groups.

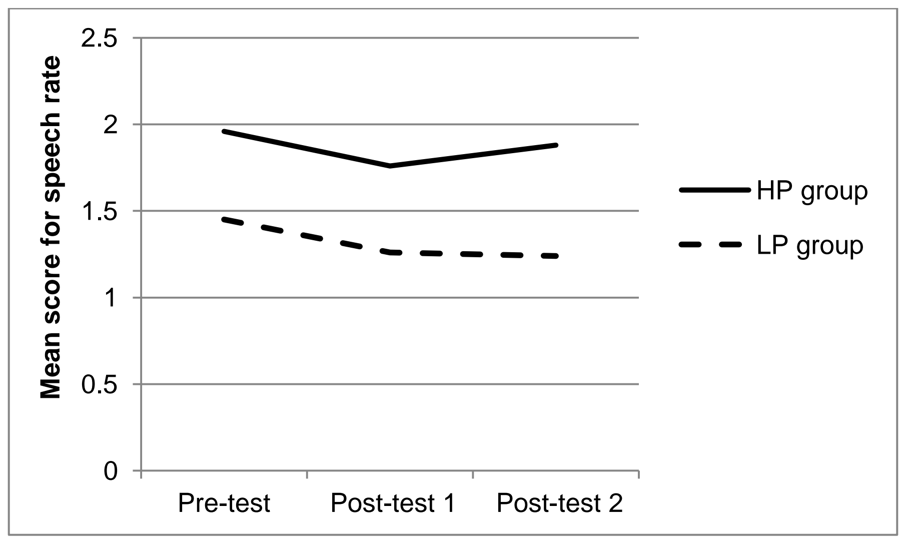


Figure 6.5.1 Line graph: Mean performance of speech rate of the two proficiency groups across time.

Table 6.5.1 Descriptive Statistics: Speech Rate (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**LP**	24	1.45 (.43)	1.78	2.62
	HP	25	1.96 (.58)	2.38	3.22
Post-test 1 (Time 2)	**LP**	24	1.26 (.33)	1.17	1.89
	HP	25	1.76 (.40)	1.66	2.76
Post-test 2 (Time 3)	**LP**	25	1.24 (.31)	1.01	1.81
	HP	25	1.88 (.44)	1.77	2.87

An independent-samples t-test showed that the difference between the groups was significant at pre-test, t (47) = -3.50, p = .001, r = .45; a large effect which accounts for 20% of the total variance. Two ANCOVAs were conducted for time 2 and 3 therefore, with time 1 as the covariate. The first ANCOVA revealed that pre-test scores were significantly related to post-test 1 scores, F (1, 45) = 40.19, $p < .001$, *partial* η^2 = .47. This was considered to be a very large effect as it explains 47% of the total variance. Having controlled for the effect of pre-

test scores, proficiency groups had a significant effect on post-test 1 scores, F (1, 45) = 7.36, $p < .01$, *partial* $\eta^2 = .14$. This was a medium effect which accounted for 14% of the total variance. In terms of post-test 2 scores, pre-test scores were also related to them, F (1, 46) = 16.76, $p < .001$, *partial* $\eta^2 = .54$. This was a very large effect as it explained 54% of the variance. After controlling for the effect of the covariate, proficiency groups significantly affected speech rate scores, F (1, 46) = 16.76, $p < .001$, *partial* $\eta^2 = .27$; a large effect which accounted for 27% of the total variance.

ii. *Mean Length of Runs*

The second variable in this construct is MLR. Graphs on Appendix 44 show the distribution of the MLR data at each time point. The K-S test indicated that the assumption of normality was met for time 3 data only: pre-test, D (48) = .17, $p = .001$; post-test 1, D (48) = .16, $p < .01$; post-test 2, D (48) = .10, $p = .20$. The results of the Levene's test showed that the assumption of homogeneity of variance was only met for time 3 data only: pre-test, F (1, 46) = 8.75, $p < .01$; post-test 1, F (1, 46) = 4.89, $p < .05$; post-test 2, F (1, 46) = .89, $p = .35$. So the results of non-parametric tests were compared to those of parametric tests on the time 1 and time 2 data, in order to determine the extent to which these agree. As shown in Appendix 46, the two types of tests yield similar results. Parametric tests therefore were used. The line graph and the table below show the descriptive statistics.

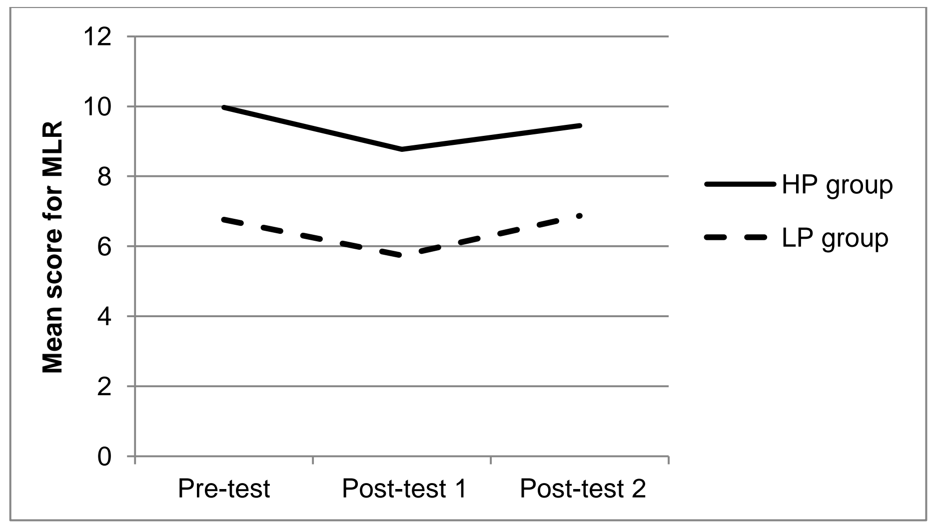


Figure 6.5.2 Line graph: Mean performance of MLR of the two proficiency groups across time.

Table 6.5.2 Descriptive Statistics: MLR (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**LP**	24	6.76 (1.90)	8.25	11.50
	HP	25	9.97 (6.36)	18	20.57
Post-test 1 (Time 2)	**LP**	24	5.74 (1.95)	9.46	10.17
	HP	25	8.77 (2.97)	11.35	15.67
Post-test 2 (Time 3)	**LP**	25	6.87 (2.06)	7.83	11.57
	HP	25	9.45 (2.86)	12.64	17

An independent-samples t-test showed a significant difference at pre-test, t (33.10) = -3.36, $p < .05$, $r = .50$. This effect accounted for 25% of the total variance, so it was considered to be a large one. An ANCOVA revealed that pre-test scores of MLR were significantly related to post-test 1 scores, F (1, 45) = 15.89, $p < .001$, *partial* $\eta^2 = .26$; a large effect which explained 26% of the variance. Having controlled for the effect of pre-test scores, proficiency groups were also significantly related to post-test 1 scores, F (1, 45) = 6.30, $p < .05$, *partial* $\eta^2 = .12$; a medium effect which explained 12% of the total variance. In terms of post-test 2 scores, pre-test scores were significantly related to them, F (1, 46) = 28.35, $p < .001$, *partial* $\eta^2 = .38$.

This was a large effect which accounted for 38% of the variance. After controlling for the effect of the covariate, the ANCOVA revealed that proficiency groups did not have a significant effect on MLR scores, F (1, 46) = 2.67, p = .11.

iii. <u>*Pause Rate*</u>

The next variable of the fluency construct is pause rate, for which the distribution graphs are presented in Appendix 44. The K-S test indicated that these were normally distributed only at time 3: pre-test, D (48) = .14, $p < .05$; post-test 1, D (48) = .18, $p = .001$; post-test 2, D (48) = .11, p = .20. The Levene's test indicated that these were homogenous only at times 1 and 3: pre-test, F (1, 46) = .55, p = .46; post-test 1, F (1, 46) = 7.23, $p < .05$; post-test 2, F (1, 46) = .001, p = .97. Comparisons of results of non-parametric to parametric tests were carried out therefore. As shown in Appendix 46, the results of both types of tests indicated a between-group difference at all time points at the same significance level. Parametric tests therefore could be used. The results of the two groups in terms of this variable are shown in Figure 6.5.3, as well as Table 6.5.3 below.

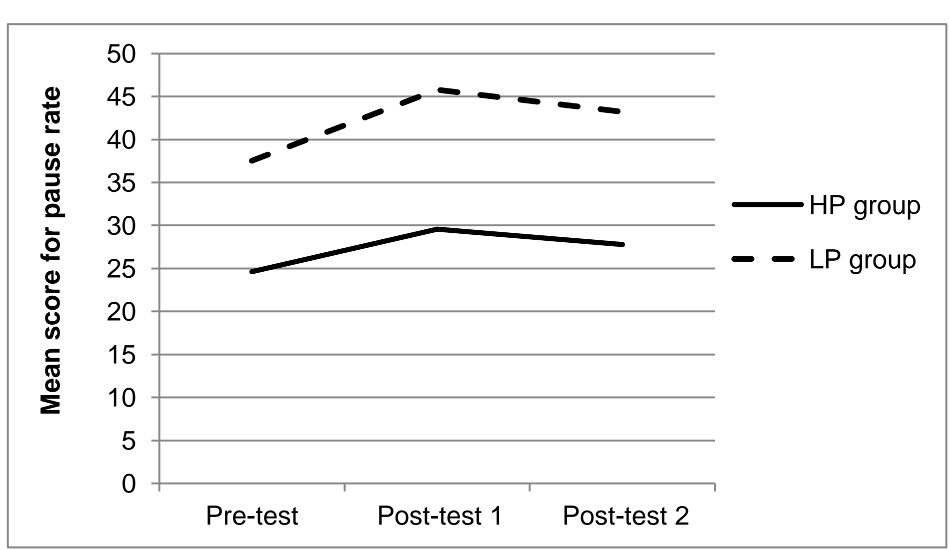


Figure 6.5.3 Line graph: Mean performance of pause rate of the two proficiency groups across time.

Table 6.5.3 Descriptive Statistics: Pause Rate (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**LP**	24	37.53 (12.45)	51.08	65.72%
	HP	25	24.61 (11.05)	48.34	59.25%
Post-test 1 (Time 2)	**LP**	24	45.79 (16.11)	60.14	84.92%
	HP	25	29.58 (9.70)	42.76	56.29%
Post-test 2 (Time 3)	**LP**	25	43.21 (10.23)	42.78	68.60%
	HP	25	27.77 (10.89)	43.02	57.44%

An independent-samples t-test showed a significant difference between the two groups at pre-test, t (47) = 3.85, p < .001, r = .49. The effect of groups was large as it accounted for 24% of the total variance. The first ANCOVA indicated that pre-test scores of pause rate were significantly related to post-test 1 scores, F (1, 45) = 35.72, p < .001, *partial* η^2 = .44; a large effect which accounted for 44% of the total variance. Having controlled for the effect of pre-test scores, the effect of proficiency groups on post-test 1 scores was approaching significance, F (1, 45) = 3.79, p = .058, *partial* η^2 = .08; a medium effect which explained 8% of the variance. The second ANCOVA revealed that pre-test scores' effects were significantly related to post-test 2 scores, F (1, 46) = 32.78, p < .001, *partial* η^2 = .42; a very large effect size as it accounted for 42% of the total variance. When controlling for the effect of the covariate, proficiency was significantly related to post-test 2 scores, F (1, 46) = 8.44, p < .01, *partial* η^2 = .15. The effect of proficiency was medium to large as it explained 15% of the variance.

The results of the temporal variables are very similar. Despite excluding initial pre-test differences, significant differences continued to occur between the two groups at post-tests on all three temporal variables. The HP group did significantly better on both post-tests in terms of speech rate; scored significantly higher than the LP group on MLR at post-test 1; and had a

significantly lower pause rate than the LP group at post-test 2. The results so far therefore seem to suggest that TR+ benefits more the fluency of HP students rather than that of LP students.

iv. *Repetitions*

The results of the three dysfluency markers follow. First, the progress of the two groups on the amount of repetitions is shown in Figure 6.5.4 below. The distribution graphs for these data appear in Appendix 45. According to the K-S test, the data were not normally distributed: pre-test, D (48) = .25, $p < .001$; post-test 1, D (48) = .23, $p < .001$; post-test 2, D (48) = .19, $p < .001$. The Levene's test indicated that the assumption of homogeneity was met for times 2 and 3 only: pre-test, F (1, 46) = .4.60, $p < .05$; post-test 1, F (1, 46) = .03, $p = .86$; post-test 2, F (1, 46) = 3.21, $p = .08$. As the two assumptions were not met for all time points, comparisons of results of non-parametric to those of parametric tests were made. Appendix 46 shows that the results of the two types of tests agree. Since there is no difference therefore, parametric tests were used. The descriptive statistics follow in Table 6.5.4.

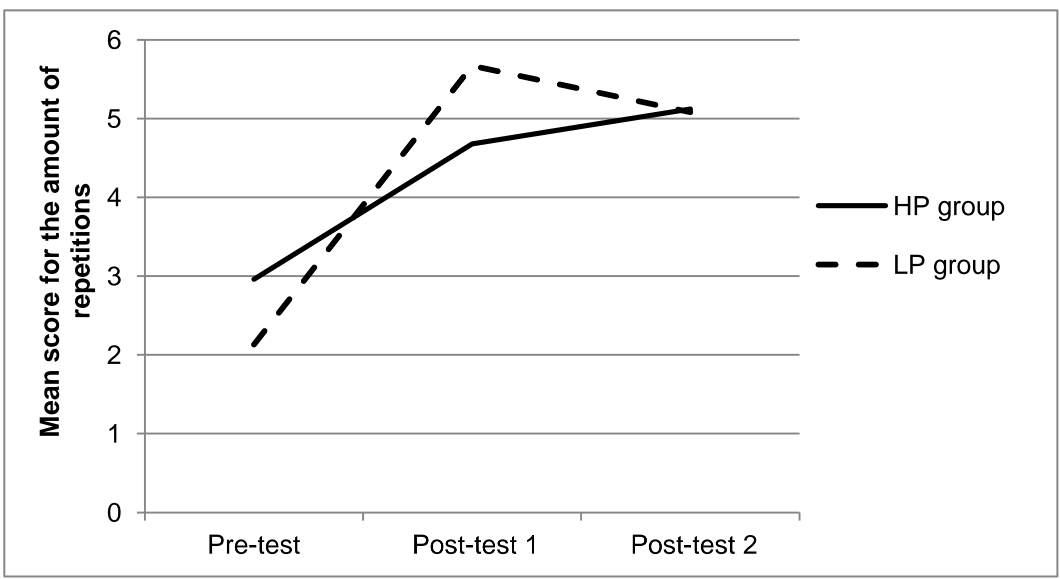


Figure 6.5.4 Line graph: Mean performance of the amount of repetitions of the two proficiency groups across time.

Table 6.5.4 Descriptive Statistics: Repetitions (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**LP**	24	2.13 (1.62)	6	6
	HP	25	2.96 (2.83)	10	10
Post-test 1 (Time 2)	**LP**	24	5.67 (6.26)	13	14
	HP	25	4.68 (6.24)	14	14
Post-test 2 (Time 3)	**LP**	25	5.08 (6.32)	18	18
	HP	25	5.12 (5.92)	18	18

The independent-samples t-test showed that the difference between the two groups was non-significant at pre-test, t (38.50) = -1.27, p = .21. A mixed ANOVA was conducted therefore. The Mauchly's test indicated that the assumption of sphericity was met, χ^2 (2) = 3.67, p = .16. A significant main effect of time on the amount of repetitions was found, F (2, 92) = 11.57, p < .001, *partial* η^2 = .20. The effect of this variable was almost large as it accounted for 20% of the variance. This suggests that the time of the test significantly affected the amount of repetitions. Contrasts revealed that the difference lay between time 1 and time 2, F (1, 46) = 23.99, p < .001, *partial* η^2 = .34 (large effect accounting for 34% of variance), but not between time 2 and time 3, F (1, 46) = .95, p = .95. Pairwise comparisons indicated that the difference was also significant between time 1 and time 3, p < .001.

As for the between-group differences, a non-significant effect was found, F (1, 46) = .01, p = .91. This suggests that the overall amount of repetitions did not differ according to groups. In terms of the interaction effect, a non-significant effect was found between the time of the test and the groups, F (2, 92) = 1.01, p = .37. This indicates that the change of the amount of repetitions across time did not differ according to groups.

v. <u>Reformulations</u>

The amount of reformulations is the next dysfluency marker. The distribution graphs for this variable appear in Appendix 45. According to the K-S test, the data were not normally distributed: pre-test, D (48) = .18, $p < .001$; post-test 1, D (48) = .20, $p < .001$; post-test 2, D (48) = .15, $p < .01$. The Levene's test indicated that the assumption of homogeneity was met at time 1 and 3: pre-test, F (1, 46) = 2.16, $p = .15$; post-test 1, F (1, 46) = 4.45, $p < .05$; post-test 2, F (1, 46) = .15, $p = .70$. Comparisons of results between non-parametric to parametric tests are shown in Appendix 46. These seem to agree apart from one time point. Since there are no equivalent non-parametric tests of mixed designs however, parametric tests were used. The graph and table below give the progress of the groups in terms of reformulations.

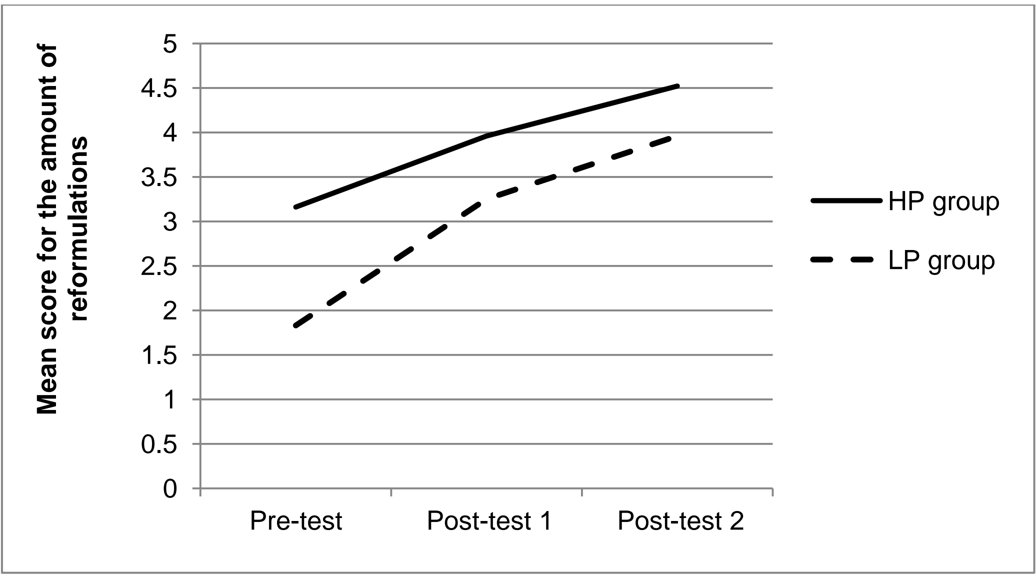


Figure 6.5.5 Line graph: Mean performance of the amount of reformulations of the two proficiency groups across time.

Table 6.5.5 Descriptive Statistics: Reformulations (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**LP**	24	1.83 (1.66)	5	5
	HP	25	3.16 (2.69)	14	14
Post-test 1 (Time 2)	**LP**	24	3.25 (2.27)	10	10
	HP	25	3.96 (3.86)	14	14
Post-test 2 (Time 3)	**LP**	25	3.96 (3.23)	10	10
	HP	25	6.52 (3.51)	14	14

The results of an independent-samples t-test showed that the difference between the two groups was significant at pre-test, t (47) = -2.07, $p < .05$, $r = .29$. According to Cohen's suggestions (1992), this was a medium effect, as it accounted for 8.4% of the total variance. In order to control for these pre-test differences, two ANCOVAs were conducted for the remaining two time points. The first ANCOVA showed that the pre-test scores were not significantly related to post-test 1 scores, F (1, 45) = 3.09, $p = .09$. Similarly, the effect of proficiency groups on post-test 1 scores was non-significant, F (1, 45) = .09, $p = .77$. The second ANCOVA revealed that pre-score effects were significantly related to post-test 2 scores, F (1, 46) = 4.81, $p < .05$, *partial* $\eta^2 = .09$. This effect was medium as it explained 9% of the total variance. After controlling for the effect of the covariate, proficiency groups were not significantly related to post-test 2 scores, F (1, 46) = .004, $p = .95$.

vi. *Incomplete words*

The final dysfluency marker is the amount of incomplete words. The distribution graphs for this variable are included in Appendix 45. The K-S test indicated that the normality assumption has not been met: pre-test, D (48) = .27, $p < .001$; post-test 1, D (48) = .18, $p < .001$; post-test 2, D (48) = .18, $p < .001$. The Levene's test indicated that the assumption of

homogeneity of variance was met: pre-test, *F* (1, 46) = .27, *p* = .61; post-test 1, *F* (1, 46) = 1.24, *p* = .27; post-test 2, *F* (1, 46) = 1.46, *p* = .23. When the results of non-parametric tests were compared to parametric tests on these data however, it was revealed that these two types of tests yield the same results. Parametric tests were used therefore. Details on the groups' progress are shown below.

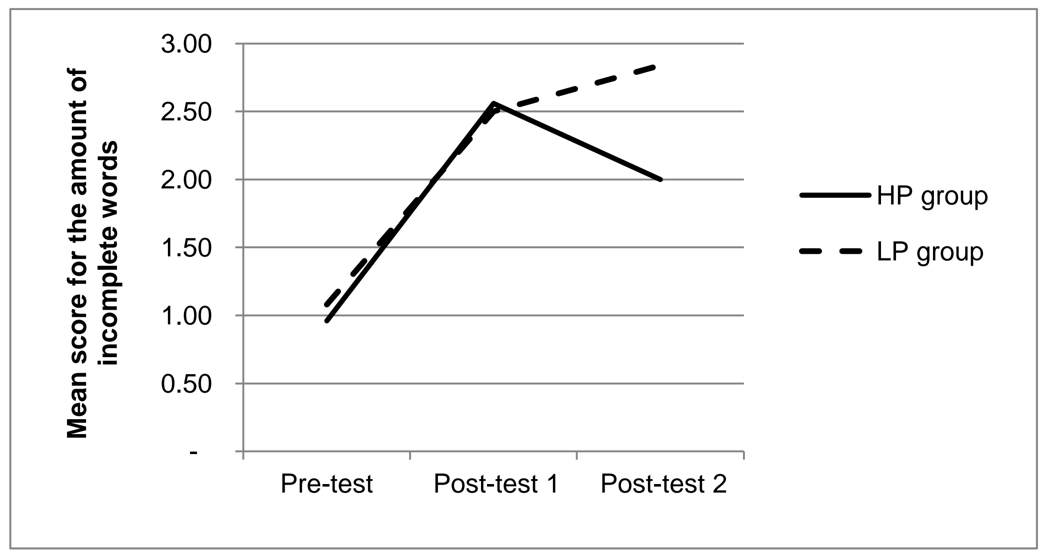


Figure 6.5.6 Line graph: Mean performance of the amount of incomplete words of the two proficiency groups across time.

Table 6.5.6 Descriptive Statistics: Incomplete words (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**LP**	24	1.08 (1.21)	4	4
	HP	25	.96 (1.34)	4	4
Post-test 1 (Time 2)	**LP**	24	2.50 (2.25)	7	7
	HP	25	2.56 (3.37)	12	12
Post-test 2 (Time 3)	**LP**	25	2.84 (2.36)	8	8
	HP	25	2.00 (1.85)	6	6

An independent-samples t-test showed that the difference between the two groups was non-significant at pre-test, *t* (47) = .34, *p* = .74. A mixed ANOVA was conducted therefore. As the Mauchly's test showed that the assumption of homogeneity of variance was violated, χ^2

(2) = 36.51, $p < .001$, the results of the Greenhouse-Geisser test are reported here. The main effect of time on the amount of incomplete words was found to be significant, F (1.74, 80.02) = 8.76, $p < .001$, *partial* $\eta^2 = .16$. This effect was medium to large as it accounted for 16% of the total variance. Contrasts revealed that this difference lay between time 1 and time 2, F (1, 46) = 13.47, $p = .001$, *partial* $\eta^2 = .23$ (large effect explaining 23% of variance), but not between time 2 and time 3, F (1, 46) = .14, $p = .71$. Pairwise comparisons indicate that the difference was also significant between time 1 and time 3, $p < .001$. As for the between-group differences, a non-significant effect was found, F (1, 46) = .68, $p = .41$. This suggests that the amount of incomplete words did not differ according to groups. Similarly, a non-significant interaction effect was found between the time of the test and the groups, F (2, 92) = .54, p = .59. This indicates that the change of the amount of incomplete words across time did not differ according to groups.

The only significant results of the dysfluency markers come from the effect of time on the amount of repetitions and incomplete words. These two significantly increased from pre-test to post-tests. The absence of significant interaction effects and group effects allows for no claims to be made based on dysfluency markers.

6.5.2 Content variables

The performance of the two groups was also compared in terms of content, which is measured based on *verb ratio, L1 syllables* and *idea units* in this study.

i. *Verb Ratio*

Looking at verb ratio first, its distribution graphs appear in Appendix 47. According to the K-S test, these data were normally distributed at times 2 and 3: pre-test, D (48) = .14, $p < .05$;

post-test 1, D (48) = .09, p = .20; post-test 2, D (48) = .09, p = .20. The Levene's test showed that the data were homogenous at all three time points: pre-test, F (1, 46) = .01, p = .92; post-test 1, F (1, 46) = .09, p = .76; post-test 2, F (1, 46) = .79, p = .38. Since normality requirements have not been met for all three time points, the results of non-parametric tests on the data were compared to those of parametric tests. As shown in Appendix 48, the results were the same. Parametric tests were conducted therefore. The descriptive results for the verb ratio are presented in the graph and the table below.

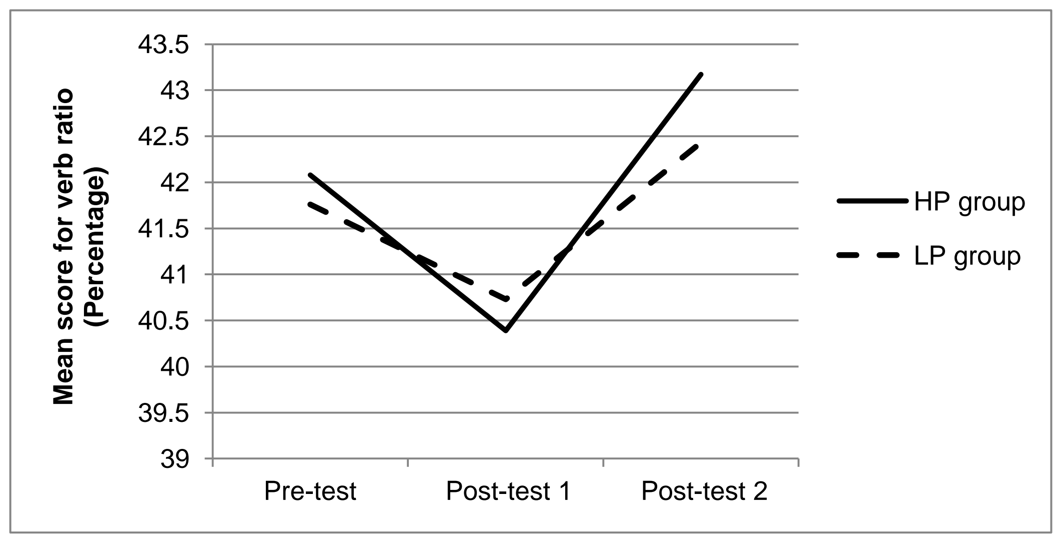


Figure 6.5.7 Line graph: Mean performance of verb ratio of the two proficiency groups across time.

Table 6.5.7 Descriptive Statistics: Verb Ratio (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**LP**	24	41.76 (11.26)	56.06	83.33%
	HP	25	42.08 (9.40)	40	69.23%
Post-test 1 (Time 2)	**LP**	24	40.73 (9.72)	39.31	60%
	HP	25	40.39 (9.41)	42.65	69.23%
Post-test 2 (Time 3)	**LP**	25	42.43 (7.27)	26.07	57.89%
	HP	25	43.17 (8.42)	35.72	62.50%

An independent-samples t-test showed no significant difference between the two groups at pre-test, t (47) = -.11, p = .91. A mixed ANOVA was carried out therefore. The Mauchly's test indicated that the assumption of sphericity has been met, χ^2 (2) = 4.73, p = .09. The effect of time was found to be non-significant, F (2, 92) = .89, $p < .05$. This suggests that the time of the test did not affect verb ratio. A non-significant main effect was also found for groups, F (1, 46) = .01, p = .94. This means that verb ratio did not differ according to groups. As for the interaction effect between time and groups, a non-significant effect was found, F (2, 92) = .05, p = .95. This indicates that the change of verb ratio across time did not differ according to groups. The fact that no significant effects were found for verb ratio was surprising, as the HP group was considered to be capable of producing a greater variety of verbs than the LP group. No claims can be made in terms of this variable therefore.

ii. *L1 syllables*

The second variable of this construct is L1 syllables. This variable shows which group tended to switch to the L1 more frequently. Graphs of the distribution of the data are presented in Appendix 47. According to the K-S test, these data were not normally distributed: pre-test, D (48) = .24, $p < .001$; post-test 1, D (48) = .25, $p < .001$; post-test 2, D (48) = .23, $p < .001$. The Levene's test showed that the data were not homogenous either: pre-test, F (1, 46) = 13.24, p = .001; post-test 1, F (1, 46) = 8.10, $p < .01$; post-test 2, F (1, 46) = 19.31, $p < .001$. Non-parametric tests and their equivalent parametric tests were conducted on the data therefore in order to determine the extent to which these match. As shown in Appendix 48, the results of the two types of tests are the same. Parametric tests were used therefore. Figure 6.5.8 shows the progress of the groups in terms of this variable, while Table 6.5.8 shows the descriptive statistics.

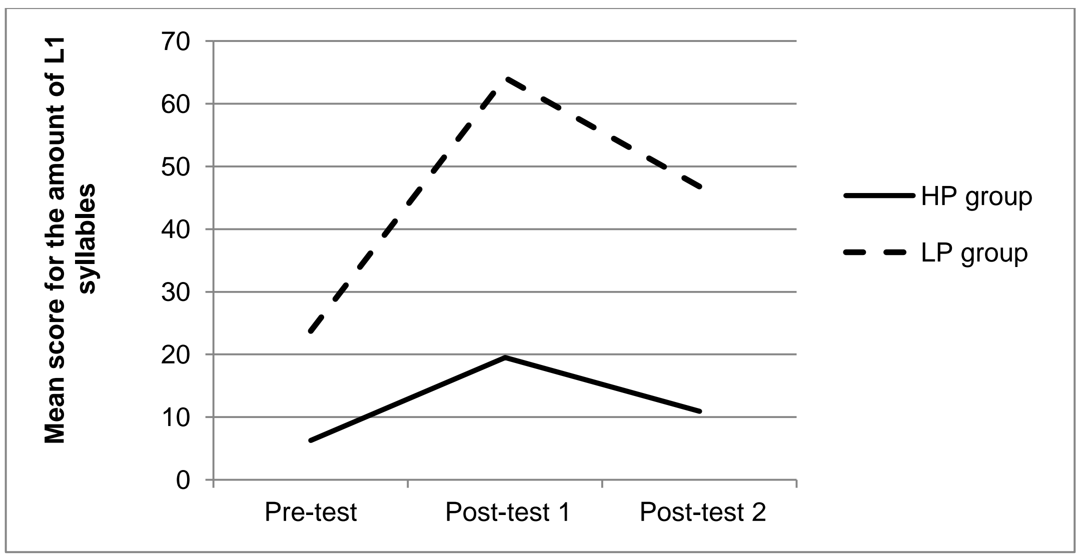


Figure 6.5.8 Line graph: Mean performance of the amount of L1 syllables of the two proficiency groups across time.

Table 6.5.8 Descriptive Statistics: L1 Syllables (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test	**LP**	25	23.72 (26.83)	90	90
	HP	25	6.28 (6.87)	77	79
Post-test 1	**LP**	24	64.13 (71.08)	269	272
	HP	25	19.48 (21.13)	77	79
Post-test 2	**LP**	25	46.76 (48.54)	180	180
	HP	25	10.92 (12.48)	51	51

An independent-samples t-test showed that the difference of the amount of L1 syllables was significant at pre-test between the two groups, t (27.13) = 3.15, $p < .05$, $r = .52$. This effect was large, as it accounted for 27% of the variance. Two ANCOVAs were conducted therefore for the two remaining time points. The first ANCOVA showed that the pre-test scores of L1 syllables were significantly related to post-test 1 scores, F (1, 46) = 25.13, $p < .001$, *partial* η^2 = .35. This was a large effect that accounted for 35% of the variance. However, once the effect of the covariate was controlled, the effect of proficiency groups became non-significant, F (1, 46) = 1.96, $p = .17$. Similarly, pre-scores were significantly related to post-

test 2 scores, F (1, 47) = 73.03, p < .001, *partial* η^2 = .61 (a very large effect explaining 61% of variance), but after controlling for this effect, proficiency groups were not significantly related to post-test 2 scores, F (1, 47) = 2.60, p = .11.

iii. *Idea units*

Finally, the two groups are compared in terms of *idea units.* The distribution graphs of the data are shown in Appendix 47. The K-S test indicated that these data were not normally distributed: pre-test, D (48) = .16, p < .01; post-test 1, D (48) = .19, p < .001; post-test 2, D (48) = .17, p < .01. As for homogeneity of variance, the Levene's test showed this assumption was only met at time 3: pre-test, F (1, 46) = 5.33, p < .05; post-test 1, F (1, 46) = 12.24, p = .001; post-test 2, F (1, 46) = 3.04, p = .09. Comparisons of non-parametric tests' results to parametric tests' results indicated that the two types of tests yielded the same results (Appendix 48). Parametric tests were conducted therefore and the descriptive results are shown below.

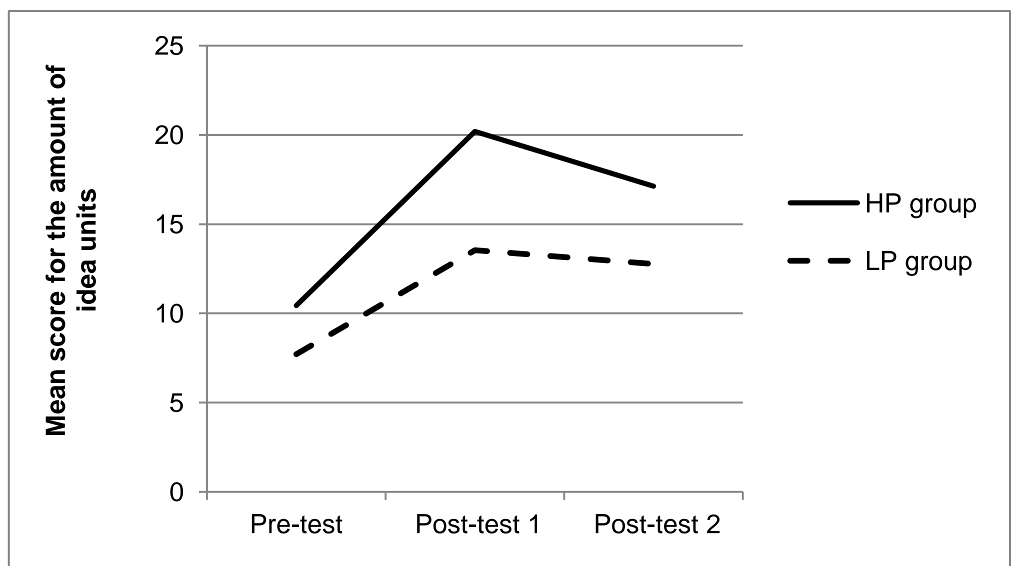


Figure 6.5.9 Line graph: Mean performance of the amount of idea units of the two proficiency groups across time.

Table 6.5.9 Descriptive statistics: Idea units (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**LP**	24	7.71 (2.18)	9	13
	HP	25	10.44 (3.35)	11	16
Post-test 1 (Time 2)	**LP**	24	13.54 (2.75)	12	20
	HP	25	20.20 (8.00)	29	41
Post-test 2 (Time 3)	**LP**	25	12.76 (3.34)	12	18
	HP	25	17.12 (7.13)	35	40

An independent-samples t-test showed that the difference between the two groups at pre-test is significant, t (41.36) = -3.39, $p < .05$, $r = .47$; a large effect which accounts for 22% of the variance. Two ANCOVAs were conducted for the two remaining time points. The first ANCOVA revealed that pre-test scores were significantly related to post-test 1 scores, F (1, 45) = 13.90, $p < .001$, *partial* $\eta^2 = .25$. This effect was large as it explained 25% of the variance. Having controlled for this effect, post-test 1 scores were also significantly affected by proficiency groups, F (1, 45) = 4.38, $p < .05$, *partial* $\eta^2 = .09$; a medium effect explaining 9% of the variance. As for post-test 2, pre-test scores were not significantly related to post-test 2 scores, F (1, 46) = 2.40, $p = .13$. Similarly, the effect of proficiency groups was non-significant, F (1, 46) = 2.98, $p = .09$.

Overall, the only significant result from the content variables is that of idea units at time 2. This indicated that, after controlling for initial pre-test differences, the HP group still produced significantly more idea units than the LP group. It could be argued therefore that TR+ benefits more HP students in terms of content in relation to LP students.

6.5.3 Accuracy variables

The two variables in this construct determine whether accuracy is maintained in case one of the proficiency groups develops fluency.

i. <u>Percentage of error-free clauses</u>

The first variable is the percentage of error-free clauses. The distribution graphs for this variable are included in Appendix 49. According to the K-S test, these data were normally distributed: pre-test, D (48) = .13, p = .17; post-test 1, D (48) = .12, p = .11; post-test 2, D (48) = .11, p = .15. The Levene's test however showed that they were homogenous only at times 1 and 3: pre-test, F (1, 46) = 3.83, p = .06; post-test 1, F (1, 46) = 6.94, $p < .05$; post-test 2, F (1, 46) = 1.61, p = .21. When conducting parametric and non-parametric tests on the data, these showed that the results were the same (Appendix 50). Parametric tests were selected for further explorations therefore. The descriptive statistics results are shown in the graph and table below.

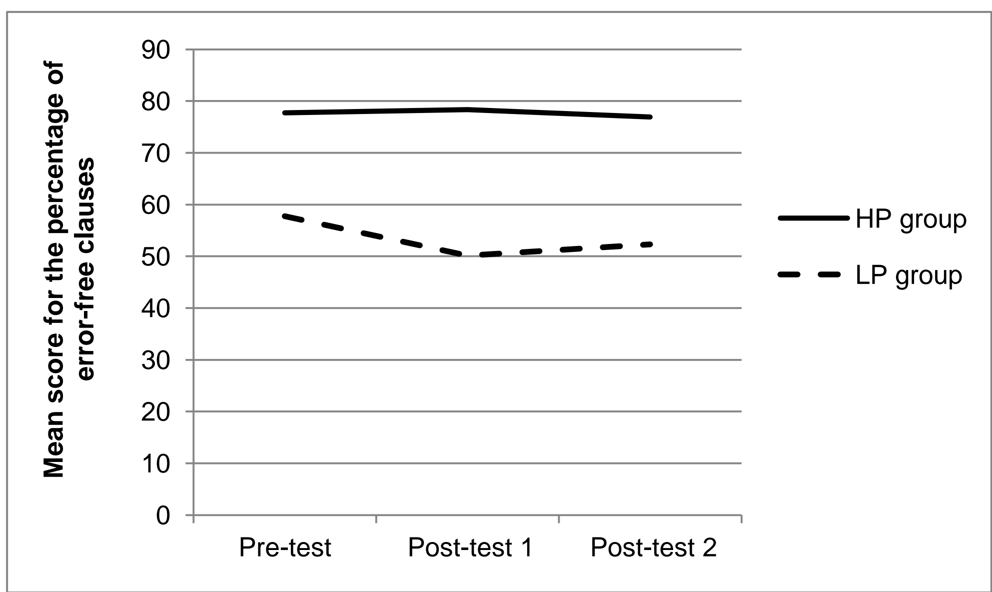


Figure 6.5.10 Line graph: Mean performance of the percentage of error-free clauses of the two proficiency groups across time.

Table 6.5.10 Descriptive Statistics: Percentage of error-free clauses (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**LP**	24	57.77 (24.13)	100	100%
	HP	25	77.75 (15.69)	63.64	100%
Post-test 1 (Time 2)	**LP**	24	50.15 (27.47)	100	100%
	HP	25	78.38 (17.10)	63.64	100%
Post-test 2 (Time 3)	**LP**	25	52.34 (25.92)	88.89	88.89%
	HP	25	76.94 (19.12)	80	100%

An independent-samples t-test showed that the difference between the two groups at pre-test was significant, t (47) = -3.45, p = .001, r = .45. This effect size was medium to large as it explained 20% of the total variance. Two ANCOVAs were conducted therefore. The first ANCOVA indicated that pre-test scores were significantly related to post-test 1 scores, F (1, 45) = 25.21, p < .001, *partial* η^2 = .36; a large effect explaining 36% of the variance. Having controlled for this effect, the two groups were still significantly different at post-test 1, F (1, 45) = 5.95, p < .05, *partial* η^2 = .12; a medium effect accounting for 12% of the variance. As for post-test 2 scores, these were significantly related to pre-test scores, F (1, 46) = 13.78, p = .001, *partial* η^2 = .23; a large effect explaining 23% of the variance. After controlling for this effect, it was found that the effect of proficiency groups was only approaching significance, F (1, 46) = 3.92, p = .054, *partial* η^2 = .08; a medium effect accounting 8% of the variance.

ii. *Percentage of correct verb use*

The second variable in the accuracy category is the percentage of correct verb use. Graphs showing the distribution of the data are shown in Appendix 49. According to the K-S test, these data were normally distributed only at time 2: pre-test, D (48) = .21, p < .001; post-test 1, D (48) = .12, p = .06; post-test 2, D (48) = .17, p = .001. The Levene's test showed that the

data were homogenous at time 3 only: pre-test, F (1, 46) = 12.22, p = .001; post-test 1, F (1, 46) = 4.55, p < .05; post-test 2, F (1, 46) = 1.35, p = .25. Again, comparisons between the results of non-parametric to parametric tests were carried out therefore. Since these yielded the same results (Appendix 50), the analysis continued with parametric tests. Details of the descriptive statistics results are shown below.

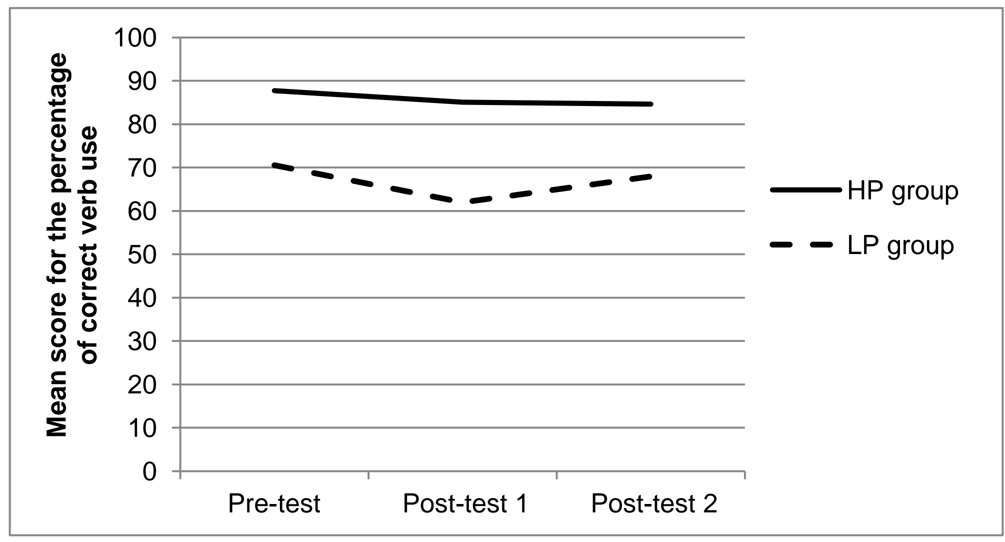


Figure 6.5.11 Line graph: Mean performance of the percentage of correct verb use of the two proficiency groups across time.

Table 6.5.11 Descriptive Statistics: Percentage of correct verb use (Proficiency groups)

	Group	N	Mean (SD)	Range	Max
Pre-test (Time 1)	**LP**	24	70.52 (26.63)	100	100%
	HP	25	87.73 (10.95)	38.46	100%
Post-test 1 (Time 2)	**LP**	24	61.94 (26.22)	100	100%
	HP	25	85.07 (13.89)	43.48	100%
Post-test 2 (Time 3)	**LP**	25	67.97 (22.50)	100	100%
	HP	25	84.66 (16.98)	80	100%

As the independent-samples t-test showed a significant difference between the two groups at pre-test, t (30.31) = -2.98, p < .05, r = .47 (a large effect explaining 22% of the variance), two

ANCOVAs were conducted for this variable. According to the first ANCOVA, pre-test scores were significantly related to post-test 1 scores, F (1, 45) = 27.85, $p < .001$, *partial* η^2 = .38; a large effect accounting for 38% f the variance. Having controlled for this effect, proficiency groups had a significant effect on post-test 1 scores, F (1, 45) = 5.23, $p < .05$, *partial* η^2 = .10; a medium effect explaining 10% of the variance. The second ANCOVA revealed that pre-test scores were significantly related to post-test 2 scores, F (1, 46) = 19.83, $p < .001$, *partial* η^2 = .30; a large effect accounting for 30% of the variance. After controlling for this effect, proficiency groups had no effect on post-test 2 scores, F (1, 46) = 1.55, p = .22.

The two accuracy variables show similar results. The only significant effect found on both measures is that of groups at time 2. This suggests that the HP group was significantly more accurate than the LP group immediately after the intervention, even when their initial differences were accounted for. TR+ therefore seems to be beneficial in terms of accuracy for the HP group more than the LP group.

6.5.4 Response to RQ5

The final research question investigated whether TR+ benefits more HP students or LP students in terms of fluency. As with the previous research question, the development of fluency is discussed in relation to content and accuracy variables. Starting with the fluency variables, the two tables below show the results of the fluency variables collectively.

Table 6.5.12 Results of temporal variables of proficiency groups

FLUENCY MEASURES (Temporal variables)		
Speech Rate	2 ANCOVAs	
	Pre-test	*sig****
	Post-test 1	*sig****
	Post-test 2	*sig****
Mean Length of Runs	2 ANCOVAs	
	Pre-test	*sig**
	Post-test 1	*sig**
	Post-test 2	*ns*
Pause rate	2 ANCOVAs	
	Pre-test	*sig****
	Post-test 1	*sig*$^{appr.}$
	Post-test 2	*sig***

sig = significant; *ns* = non-significant; $p < .05$*; $p < .01$**; $p < .001$***; appr. = approaching significance

Table 6.5.13 Results of dysfluency markers of proficiency groups

FLUENCY MEASURES (Dysfluency markers)		
Repetitions	Mixed ANOVA	
	Time	*sig**** 1 & 2*** 1 & 3***
	Groups	*ns*
	Time*Groups	*ns*
Reformulations	2 ANCOVAs	
	Pre-test	*sig**
	Post-test 1	*ns*
	Post-test 2	*ns*
Incomplete words	Mixed ANOVA	
	Time	*sig**** 1 & 2*** 1 & 3***
	Groups	*ns*
	Time*Groups	*ns*

sig = significant; *ns* = non-significant; $p < .05$*; $p < .01$**; $p < .001$***; appr. = approaching significance

As the analysis of the dysfluency markers yielded no significant results, the conclusions on fluency is based on the results of the temporal variables. Having controlled for the initial pre-test differences, these three variables suggest that overall the HP students did significantly better than the LP group. They gained a significantly higher score on speech rate at both post-tests; they had a significantly higher MLR than the LP group at post-test 1, and they had a significantly lower pause rate at post-test 2. These results support the argument that TR+ benefits HP students more than LP students in terms of temporal fluency.

Looking at the content variables next, the following table presents the results of these variables collectively.

Table 6.5.14 Results of content variables of proficiency groups

CONTENT VARIABLES		
Verb Ratio	Mixed ANOVA	
	Time	*ns*
	Groups	*ns*
	Time*Groups	*ns*
L1 syllables	2 ANCOVAs	
	Pre-test	*sig**
	Post-test 1	*ns*
	Post-test 2	*ns*
Idea units	2 ANCOVAs	
	Pre-test	*sig**
	Post-test 1	*sig**
	Post-test 2	*ns*

sig = significant; *ns* = non-significant; $p < .05$*; $p < .01$**; $p < .001$***; appr. = approaching significance

As it is evident from the table above, the only variable showing significant differences between the two proficiency groups is that of idea units. After controlling for initial pre-test differences, the HP group produced significantly more idea units immediately after the intervention than the LP group. It could be argued therefore, that based on idea units only, TR+ seems to be more beneficial for HP students than for LP students in terms of content.

Finally, the table below presents the results of the two accuracy variables collectively.

Table 6.5.15 Results of accuracy variables of proficiency groups

ACCURACY VARIABLES		
Percentage of error-free clauses	2 ANCOVAs	
	Pre-test	*sig****
	Post-test 1	*sig***
	Post-test 2	*sig*$^{appr.}$
Percentage of correct verb use	2 ANCOVAs	
	Pre-test	*sig**
	Post-test 1	*sig**
	Post-test 2	*ns*

sig = significant; *ns* = non-significant; *p* < .05*; *p* < .01**; *p* < .001***; appr. = approaching significance

According to both accuracy variables, the HP group did significantly better in terms of accuracy than the LP group immediately after the intervention. TR+ therefore seems to be more beneficial for HP students in relation to LP students in terms of accuracy.

In response to the fifth research question thus, it seems safe to argue that TR+ benefits the HP group more than the LP group in terms of temporal fluency, idea units and accuracy.

6.6. Chapter Summary

This chapter presented the quantitative analysis conducted for the last three research questions of the present thesis. In order to present the results for each variable, graphs, as well as the results of descriptive and inferential statistics were discussed.

The third research question inquired about the effects of TR+ on WTC in relation to no task repetition. The non-significant differences found between the two experimental groups and the comparison group for speaking time and L2 syllables suggested that TR+ did not lead to more WTC, than no task repetition. This research question further inquired about the effects of the two pedagogical packages on WTC. The significant difference of L2 syllables found between the CS and the EO group at time 2 indicated that task repetition with CS *plus* led to more WTC than task repetition with EO *plus.* Despite the two inquiries of RQ3 however, the fact that the CS group produced significantly more L2 syllables than both the other two groups was important.

The fourth research question firstly inquired about the effects of TR+ on fluency, in relation to no task repetition. While the results of five out of the six fluency measures were non-significant, those of the MLR measure suggested otherwise. The two experimental groups had a significantly higher MLR rate than the comparison group on both post-tests. Based on this result, it was proposed that TR+ led to more fluency than no task repetition. The differences between experimental and comparison groups were also explored in terms of content and accuracy, in order to ensure that any change in fluency would have no detrimental effect on these two variables. While no significant results were found for content, the results of the accuracy variables suggested that TR+ did not undermine accuracy.

This research question further inquired about the effects of the two pedagogical packages on fluency. The non-significant results between the CS and the EO group on either of the

fluency, content or accuracy measures, did not allow for any conclusions to be reached on this inquiry. However, two observations could be made on these results. First, the non-significant differences found for the L1 syllables variable suggested that the incorporation of codeswitching in the CS group did not lead to excessive L1 use. Second, the feedback on accuracy provided to the EO group did not seem to increase the level of accuracy of the group, as this group did not end up with significantly higher accuracy rates than the other two groups.

Finally, the fifth research question inquired about the role of proficiency in relation to the intervention. More specifically, it examined whether TR+ benefitted high proficiency students more than low proficiency students. Having examined the differences between the two proficiency groups in terms of all three variable constructs (i.e. fluency, content, accuracy), it was found that higher proficiency students benefitted more from TR+ than lower proficiency students in regards to temporal fluency, idea units and accuracy.

The above conclusions lend support to the argument put forward at the end of the literature review chapter, suggesting that the different aspects of speech should be looked at as a whole. The results reported in this chapter provided evidence that the aspects of speech studied here (WTC, fluency, content, accuracy) were interconnected and they seemed to constantly interact with each other. A change in WTC could signal a change in fluency and in turn a change in accuracy. Research that only looks at isolated features of speech therefore could omit essential developments of other variables that could be linked to those under observation. It is essential not to investigate just isolated variables therefore, as they seem to complement one another. These results are further discussed in relation to the results of past research in the Discussion chapter (Chapter 7).

CHAPTER 7: Discussion

7.1. Introduction

This chapter discusses the results of both the qualitative and the quantitative analysis of the present study in relation to the results reported in previous research. The chapter is divided into five main sections; each one representing a research question. For each research question, the extent to which the results of this study agree with those of previous research is determined.

7.2. Discussion on RQ1

The first research question aimed to provide an in-depth introduction to the context of the study by inquiring about the teachers' views on four main topics: task implementation, task repetition, metalanguage and the role of proficiency in pair work. A qualitative analysis of nineteen interviews with teachers yielded several enlightening results.

7.2.1. Task implementation

Looking at task implementation first, it was found that in general teachers were in favour of using such communicative activities because of the many different types of benefits they offered, including practical, pedagogical, cognitive and psychological benefits. Practical benefits referred to the fact that students enjoyed completing such activities because they found them very interesting. In terms of the pedagogical benefits, teachers in general acknowledged the importance of developing the learners' oral skills and tasks were one way towards such a development. Moreover, teachers referred to forming appropriate structures and finding appropriate vocabulary as examples of the cognitive challenge learners face, when completing tasks. Finally, psychological benefits were related to less extroverted

learners who were often of a lower proficiency level. According to the teachers, making a small contribution to the completion of a task or rehearsing their L2 output in front of more proficient peers, before presenting it in front of the class and the teacher, boosts the confidence of these students.

These results can be compared to the studies conducted by Carless and colleagues (Carless & Gordon, 1997; Carless, 2002, 2003, 2004, 2008) on the teachers' opinions of task implementation in Hong Kong. Like the teachers in the present study, teachers in Carless and Gordon's study (1997) reported overall positive attitudes towards task-based learning. In the same study, the practical benefit of the students enjoying tasks and finding them interesting was also reported. In terms of the pedagogical benefit however, the two groups of teachers seem to disagree. In the Cantonese context, the majority of the teachers thought that tasks could have some benefit in language learning, but this benefit would not be greater than that gained by any other non-communicative activities. However, teachers in Cyprus stressed the importance of the positive effect that tasks can have on oral production, with one teacher arguing that tasks should be used more frequently in language teaching. In general, teachers' in Cyprus appeared to place a lot of emphasis on the development of their students' oral skills; perhaps due to the language learning policies in this context stressing the importance of oral communication. No reference was made to the cognitive or the psychological benefits of tasks in the Cantonese context.

To sum up the discussion so far therefore, it could be argued that the results of the present study are partly compatible with those of previous research. While teachers in both Cyprus and Honk Kong are in favour of task-based teaching, teachers in Cyprus find that tasks serve the purpose of developing oral skills, while teachers in Hong Kong do not see a unique benefit that can be gained from this approach.

Continuing on task implementation, both groups of teachers referred to the difficulties they face when using tasks in their classrooms. Looking at the responses of the teachers in Cyprus first, the difficulties of task implementation seem to be linked with one another. Classrooms in primary schools contain mixed-proficiency students because the majority of them attend private lessons of English and therefore their level varies. Nevertheless, the policy of schools, as well as that of the Ministry of Education and Culture, maintains that all students, despite their proficiency level, should be able to follow the lesson at school. In order for lower proficiency students to participate in the lesson, the books contain activities that are very much guided; these can hardly be considered communicative. If teachers' wish for their class to use more open-ended activities, that are much closer to the definition of tasks, they are the ones responsible for finding this new material and distributing it to their students. This responsibility led the teachers to refer to three more problems. These include the lack of time for students to complete tasks during the lesson, as well as the lack of time for teachers to prepare the material; also, the teachers' lack of confidence when teaching the subject because of their lack of expertise; and finally, their need for more innovative ideas that would make finding material faster through teacher training courses or through a bank of material. It is clear therefore that the teachers in this context reported several problems that they face with task implementation, which can be linked in some way.

Similarly, in Carless' studies, teachers in Hong Kong reported having several problems with task implementation, which are mentioned again later on in this chapter. However, only one of the problems they faced has been reported in the Cypriot context as well; that is lack of time. As in Cyprus, teachers in Hong Kong reported lack of time as a problem with both preparing and executing tasks (Carless & Gordon, 1997; Carless, 2002). This was also a finding reported by Li (1998), who argued that students in the Korean context are aimed towards passing exams and therefore there is no time for completing tasks. It could be argued

therefore that in general teachers view tasks as a time-consuming activity, rather than something that directly contributes to language learning. This belief shows therefore that tasks are not favoured when students have to be prepared for exams, as in the Cantonese context, or when teachers feel the pressure of covering a certain curriculum, as in the Cypriot context. Based on this distinction, a debate between “real” leaning and preparing for exams arises.

Many reasons could contribute to the teachers’ belief of tasks being time-consuming activities. Firstly, the instructions of tasks should be clear to all students in a class in order for tasks to be effective. Teachers therefore might dedicate time for clarifying instructions in order to ensure that all students have understood what they should do. Secondly, studies that examined pair or group work in the past have found that parts of the language used by students, when engaged in pair or group work, is off-task talk. Alley (2005) in his study of eighteen English-speaking high school learners of Spanish reported finding that 21% of the total talk was off-task talk. Instances of off-task talk were also found in Antón and DiCamilla’s (1999) study in the interactions between English-speaking high schools learners of Spanish. Carless (2002) in his studies in Hong Kong identified off-task talk used by the Cantonese learners. This off-task discussion takes up time from task completion and could therefore contribute to the perception of tasks being time-consuming activities. Finally, the fact that students should use metalanguage, either in their L1 or in their L2, in order to reach intersubjectivity often requires time. Especially in contexts where foreign languages are considered as secondary subjects and therefore less time is dedicated to them, it seems that it is not easy for teachers to use communicative activities very often, when a curriculum should be covered.

Problems with mixed-proficiency classes, inadequacy of materials, lack of teacher expertise and the need for more innovative ideas were not reported in other studies of teachers’

opinions on task implementation. However, teachers in Carless' studies reported facing other kinds of problems. One of the main problems was that students regarded pair work as an opportunity to switch to their mother tongue (Carless & Gordon, 1997; Carless, 2002, 2007). This appeared to be of particular importance to Cantonese teachers because it was reported as a main problem in all the above studies and because teachers' said that they regarded this as their own 'failure'.

Teachers in Cyprus did not report L1 use being a problem in task-based teaching. Linking this perception with the discussion we had in the interviews about metalanguage, it emerged that teachers regard the use of L1 for metalanguage as a 'natural' thing. Talking about the potential of using metalanguage in the L2, it emerged that at the moment, based on the current foreign language educational system, students should not be expected, or required, to use the L2 for metalanguage. They implied therefore that this kind of language, which almost always takes place when completing tasks, would occur in the L1. Teachers referred to the new curriculum, that was implemented the academic year following the present study, as a way towards the use of metalanguage in the L2, and therefore the reduction of L1 use. Regarding the L1 use being reported as a problem of task implementation in previous research therefore, it seems that teachers in the present study do not expect their students to be able to remain in the L2 for metalanguage and this is probably why they did not report this as a problem.

Another problem reported in Carless' studies, but not in the present study, was the lack of discipline that group work could often cause (Carless, 2002; 2003). In the present study, teachers referred to indiscipline when they were asked about the role of interlocutor familiarity. Most of them replied that they do not let their students choose their interlocutor mainly because of the fuss caused with moving around the classroom. This was not reported however, as a problem of group work. An explanation could be the excitement and interest

the students show when they complete tasks. Most of the teachers in the present study reported that their students enjoy completing tasks very much and demonstrate a particular interest. Perhaps in the case of the Cypriot context, this interest is directed towards completing the activity, rather than disturbing classroom discipline.

Concluding on the task implementation section, it was found that teachers in both the Cypriot and the Cantonese context have positive attitudes towards tasks, but they disagree in terms of the pedagogical value they see in this approach. While in previous studies teachers believe that tasks offer a benefit, but not a larger one than that offered by any other types of activities, teachers in the present study saw an invaluable benefit of tasks; that of the development of oral communication skills. In addition, both groups of teachers agreed that task implementation does not come without problems or difficulties. While both groups agree on the consumption of time as a problem with preparing and executing tasks, the two groups reported different kinds of problems. Cypriot teachers referred to mixed-proficiency groups, inadequacy of material, lack of teacher expertise and need for innovative ideas, and Cantonese teachers referred to the students' use of the L1 and the lack of discipline caused by working in pairs or groups.

7.2.2. Task repetition

Task repetition was the second topic of discussion in the interviews with the teachers. As, to my knowledge, there are no studies on the opinions of teachers on task repetition, the results of the present study are discussed here in relation to those of empirical studies on task repetition. It emerged that overall teachers prefer to conduct repetitions with types of tasks that have previously been practised, but on a new topic, rather than with exactly the same task. The reason quoted by all teachers is because students lose interest when repeating

exactly the same activities. Benefits teachers anticipate from repeating the same type of task included building on previous knowledge, having the opportunity to focus more on language and having the chance to go through things they have not comprehended well on the previous occasion.

Nevertheless, two teachers reported two different occasions where they might repeat exactly the same tasks in their classrooms. The first occasion reported by Alison is when the task is a game that the students enjoyed a lot. However, she did not think that this repetition could have any real contribution to language learning. The only benefit she reported was a psychological one for lower proficiency students, who feel that they contribute something and that they become part of the team. The second occasion where same-task repetition was reported being used is when this is initially conducted at the end of the one lesson and then repeated at the beginning of the following lesson as a continuation of what has been done.

None of the teachers mentioned potential benefits of task repetition specifically in terms of fluency, accuracy and complexity. Studies that looked at the effects of task repetition directed their focus towards these three components of oral production. Bygate (2001) found that being exposed to the same type of task for a long period of time had a significant positive effect on fluency. Since he found no evidence suggesting that this effect was transferred to a new task, he could not conclude that task repetition contributed to acquisition. Ahmadian (2011) however, in a similar type of study, was able to build on Bygate's (2001) study when he found that massed repetition led to fluency and complexity benefits that were transferable onto a new type of task.

As discussed in the literature review chapter however, these studies did not come without problems. Methodological problems included the design of the studies which seemed to have favoured experimental conditions, as well as the rather simplistic way complex concepts,

such as fluency, accuracy and complexity, were measured. Nevertheless, through this comparison of the outcomes of task repetition, it is worth noting that previous studies on task repetition examined specific aspects of oral production in order to determine the benefits of task repetition, while in the present study teachers' reports were not so specific.

Considering this idea of task repetition however, and especially 'massed' task repetition as it was examined and found effective in previous research, its ecological validity should be re-considered. How feasible is it for teachers, as well as learners, to repeat the same task repeatedly over a period of time? And how possible is it for learners to maintain the same level of interest in what they are doing on all these occasions? Especially in classroom contexts where the main purpose of language learning is to pass exams, rather than actually learn the language, the idea of task repetition would probably not be favoured. As the teachers in the present study maintained, repeating the same type of task on different topics is a more viable practice.

7.2.3. Metalanguage

The third topic of discussion during the teacher interviews concerned metalanguage. All teachers stated that this kind of language always occurs in the L1 mainly for three reasons: first, time constraints which do not allow learners to try to express metalanguage in the L2, as this would be too time-consuming; second, the relatively high level of proficiency required for L2 metalanguage, which is not always encountered with mixed-proficiency groups; and third, the relationship between learners, who have to feel comfortable with each other and not to feel embarrassed in front of their peers. Nonetheless, as it was briefly mentioned in the task implementation section earlier in this chapter, teachers seem to be understanding of the fact that their students use the L1 for metalanguage. Based on the experience of language learning

that these students had so far, teachers did not expect them to be able to maintain the L2 for metalanguage. It could even be said that they thought it was 'natural' for their students to use the L1 for this purpose, as none of them reported intervening when they heard their students using their L1. Lastly, the teachers expressed their belief that with the new curriculum, implemented on the following academic year, the goal for L2 metalanguage was more likely to be achieved.

As with task repetition, to my knowledge no studies have been conducted on the opinions of teachers on their learners' use of metalanguage. However, empirical studies that used Vygotsky's sociocultural theory as a framework for the analysis of learner codeswitching, agreed with the teachers in the present study in the sense that they found the use of L1 for metalanguage as a 'natural' process. Some of the most characteristic conclusions in this area come from Brooks and Donato (1994) who claimed that in their study the L1 was used as an important tool for '*initiating* and *sustaining* further discourse' (266), and from Antón and DiCamilla (1999), who characterised it 'as a critical psychological tool that enables learners to construct effective collaborative dialogue [...] (245). It can be concluded therefore that the teachers' acceptance of the L1 in the interactions of their learners' corresponds to the acceptance of the L1 as a facilitative tool by researchers.

This is the idea on which the present study has attempted to build on. The pattern of codeswitching *plus* proposed the recycling of this L1 use into the L2. The aim of tasks, and of language learning in general, is to learn how to communicate effectively in the L2. Research therefore should focus on how this facilitative L1 use could truly contribute to language acquisition; not just simply accept its existence. By taking advantage of this functional L1 use and turning it into L2 knowledge, codeswitching may eventually lead to acquisition, rather than just having a facilitative task role.

7.2.4. Proficiency pairing

The final topic of discussion in the interviews concerned the effects of different types of proficiency pairing on the task performance of the learners. Teachers in the present study reported that overall they were in favour of mixed-proficiency pairs, as it was supported by Iwashita (2001), because this is the kind of pairing that facilitates scaffolding. Although teachers did not actually use the term 'scaffolding', it was clear in their explanations that what they were referring to was the concept of scaffolded help that emerged in the analysis of previous studies on learner codeswitching. More specifically, the teachers talked about the help that lower proficiency students can get from their higher proficiency peers, through corrections of errors, finding L2 vocabulary and building self-confidence.

Indeed, the qualitative analysis of students' interactions has confirmed that scaffolded help does take place in this context when mixed-proficiency pairs work together. Although this result was a part of the analysis in chapter 5, it is worth mentioning it here in relation to the teachers' opinions on scaffolding. According to the analysis, there were two ways in which scaffolded help occurred. The first way was the *direct way*, according to which one student, typically the more proficient student, told their interlocutor the exact phrase of task language that they should use next in order to complete the task. The second way of offering scaffolded help was the *indirect way,* according to which help is offered through prompts. In the present study, this type of help took the form of signalling errors, offering the opportunity to reconsider output and asking and providing L2 vocabulary.

The latter type of scaffolded help, that is related to L2 vocabulary, has also been identified in previous studies. In a study with four Japanese high school learners of English, Leeming (2011) found that the learners helped each other during tasks by finding appropriate English vocabulary collaboratively. Additionally, they used Japanese for discussing linguistic items.

Similarly, in de la Colina and Mayo's (2009) study, Spanish learners of English used their L1 for vocabulary. Nevertheless, they also used it for clarifying and discussing content, understanding information, discussing linguistic forms and grammar; types of scaffolded help that were not identified in the data of the present study. First language use for L2 vocabulary was also found in Antón and DiCamilla's (1999) study on English-speaking learners of Spanish. Apart from L2 vocabulary, the L1 was used for constructing appropriate linguistic forms, as well as evaluating and understanding the meaning of a text. It is concluded therefore that the only type of scaffolded help that was common between the data of the present study and that of previous research is offering help for findings L2 vocabulary.

Having expressed their support for mixed-proficiency pairs, some teachers referred to some pre-requisites in order for mixed-proficiency pairs to work effectively. These included the character of the students, a small proficiency gap, additional targets for high proficiency pairs, the teachers' supervision and a good relationship between classmates. Since I have not come across studies that look into the teachers' opinions on proficiency pairing, these additional requirements for effective mixed-proficiency interactions cannot be compared to previous findings.

Finally, in terms of interlocutor familiarity, teachers did not seem to be in favour of allowing their students to choose their own interlocutor due to both practical and pedagogical reasons. For practical reasons, the students always conduct tasks with the person sitting next to them. Nevertheless, the teachers did not believe that their students would benefit more if they conducted tasks with a more familiar person. On the contrary, they believed that this would lead to more off-task conversations. Their opinion contradicts with the limited evidence in research suggesting otherwise. Both O'Sullivan (2002) and Hancock (1997) found in their studies that learners who had a familiar interlocutor performed better than learners who had an unfamiliar interlocutor.

7.2.5. Conclusions on RQ1

Concluding on the discussion on the first research question, it seems that teachers in different contexts seem to have positive attitudes towards task implementation, but they disagree to the extent to which they believe that these can be beneficial for language learning. It is also clear that task implementation comes with different kinds of difficulties that might differ across contexts.

Regarding task repetition, teachers in the present study reported conducting repetitions of the same-task type, rather than exactly the same task, not only because their students tended to lose interest in same-task repetition, but also because they did not believe that this could have a real contribution to language learning. None of the teachers referred to benefits in terms of fluency, accuracy and complexity, as it has been examined in past research.

Moreover, in terms of metalanguage, teachers in the present study seem to agree with previous studies arguing that the use of L1 metalanguage is a ‘natural’ process. Their agreement is evident through the fact that they do not oppose it, like teachers in other contexts. Rather they gave their own reasons as to why metalanguage always occurs in Greek and referred to the new curriculum as a way towards the maintenance of the L2 for metalanguage.

Finally, most teachers expressed their preference for mixed-proficiency pairs, as a way of facilitating scaffolding. This comes in agreement with the concept of scaffolding of previous research on task performance. Data coming from the qualitative analysis of learner interactions supported the teachers’ reports by supporting that scaffolding does take place in this context. In terms of interlocutor familiarity, teachers in this context seemed to disagree with previous findings suggesting that being familiar with one’s interlocutor could lead to a better performance.

7.3. Discussion on RQ2

The second research question prompted a qualitative analysis of the students' interactions by focusing on any change of the content of their output from session 1 (S1) to session 2 (S2). Task language and L2 metalanguage were chosen as forms of language that would determine whether the content of the output would be enriched in some way when the task was repeated with feedback.

7.3.1. Task language

Considering task language first, examples of both increasing and decreasing richness of content from S1 to S2 were found in the transcripts of both experimental groups. While increased creativity was not a surprising result as the students were completing the tasks for the second time after receiving feedback, the reasons behind decreased creativity were sought. It was suggested that the students might have lost interest when they completed the same task (despite the fact that on S2 they received feedback) and therefore experimented less with the language.

Comparing these results with previous research is not possible, as no study to date has qualitatively looked at the change of content when a task was repeated with feedback. Studies on the effects of task repetition examined quantitative changes of fluency, accuracy and complexity (e.g. Bygate, 2001; Ahmadian, 2011). Comparisons with the results of such studies are made later on in this chapter in relation to the quantitative results of the present study.

7.3.2. L2 metalanguage

Before looking at the way L2 metalanguage evolved from S1 to S2, its functions in relation to L1 metalanguage were explored. It was found that metalanguage in both languages seemed to be serving common functions, including signalling the beginning and the end of a task, assigning roles, achieving intersubjectivity and checking the status of the task. This finding supported the part of the hypothesis of the present study suggesting that L2 metalanguage can provide the same cognitive support as L1 metalanguage, but by simultaneously meeting the pure purposes of task-based learning which is effective communication in the L2.

The only study considering the functions of L1 as well as L2 use of learners during task completion was that of DiCamilla and Antón (2012). Comparing the functions of L1 and L2 use of two age groups, the authors found that the functions that beginners used in the L1 were the same functions that advanced learners used in their L2. In other words, the same functions were served by both languages, but this differed according to age. As this is the only study looking at both the functions of the learners' L1 and L2 use, it can be said that the findings of the present study agree with those of previous research in this respect.

In terms of the inquiry of RQ2, the evolution of L2 metalanguage from S1 to S2 differed according to conditions. The CS group enriched its L2 metalanguage on S2 by using the feedback from the handout, not only as it is, but in variations. This is a very important finding because it suggests that these learners were in control of the language. Additionally, this group created their own metalanguage in the L2, which did not appear on any of the handouts. The EO group however did not show evidence of enriched L2 metalanguage in S2. In fact, it was reported that the metalanguage of this group was often reduced on S2.

Based on this result therefore, it was argued that the feedback component played a very important part in the evolution of L2 metalanguage. The CS group who received feedback on

L2 metalanguage was able to experiment and build on their L2 metalanguage. The EO group however, was not able to demonstrate the same performance possibly because they did not have the same foundation. It was argued therefore, that young learners were able to maintain the L2 for metalanguage, as long as they received adequate feedback and practice. The idea of using L2 metalanguage in this context is not as far-fetched as the teachers in their interviews perceived it to be. Perhaps teachers in this context should believe more in the skills of their students and attempt to challenge them more.

Similarly, no previous study looked at the evolution of L2 metalanguage qualitatively after task repetition and receipt of feedback. It is hard therefore to draw any comparisons with previous results.

7.4. Discussion on RQ3

The effects of the intervention on willingness to communicate (WTC) were investigated for the third research question. With speaking time and L2 syllables as the two measures of WTC, a quantitative analysis across pre-test, post-test and delayed post-test yielded relevant results. As explained in previous chapters, WTC is perceived as an expansion of output, rather than an increase of turn taking.

In terms of the first inquiry comparing TR+ to no task repetition, that is comparing the performance of the experimental groups to that of the comparison group, it was found that a significant difference only existed between the CS and the COM group in terms of L2 syllables. As a significant difference between the EO and the COM group was not found, the argument in favour of TR+ could not be supported. In addition, the second enquiry of this research question prompted a comparison between the two experimental groups in terms of

WTC. The measure of L2 syllables suggested that task repetition with CS *plus* led to more WTC than task repetition with EO *plus*.

However, before interpreting the significantly higher WTC that the CS group showed in comparison to the EO group, the initial difference of proficiency between the two experimental groups should be considered. The results of the Kruskal-Wallis test in Chapter 6 have shown that the CS group scored significantly higher on the C-test than the EO group. The possibility of this difference in WTC being the result of different proficiency level should be considered.

Nevertheless, the fact that the CS group did significantly better in terms of L2 syllables than both the other two conditions is important. This supports the part of the hypothesis suggesting that the incorporation of the L1 would increase WTC. The potential relationship between codeswitching and WTC was initially proposed in a small-scale study conducted by Macaro *et al.* (2012), which was discussed in detail in the Literature Review chapter. This was the study in which a 'bilingual' group, who had a bilingual assistance and who was encouraged to codeswitch, showed a large improvement of fluency and overall speaking skills in relation to two other groups, who had a monolingual teacher and who were discouraged from using their L1. Although WTC was not empirically examined in this study, the idea that it could be enhanced with the incorporation of codeswitching was put forward. It could be argued therefore that the results of WTC in the present study build on the insights of the Macaro *et al.* (2012) paper, in the sense that the present study provides evidence of the CS conditions actually leading to higher WTC than the other two conditions.

7.5. Discussion on RQ4

The fourth research question inquired about the effects of the intervention on fluency, but by also taking into account the effects this could have on accuracy and content. More specifically, the research question assumed no detrimental effects of accuracy, in case of a potential fluency development. The quantitative analysis of pre- and post-tests presented in Chapter 6 yielded relevant results for this inquiry.

In terms of fluency first, six measures were used: speech rate, MLR, pause rate, repetitions, reformulations and incomplete words. At first glance, some of these measures suggested a decline of fluency, based on the visual decline of speech rate and MLR. This decline should not be surprising for the EO group, as the English-only conditions were expected to block their performance and therefore not allow any fluency development. Similarly, no improvement of fluency was predicted for the COM group, who only completed the tasks once. A possible decrease of fluency would be surprising for the CS group however because the hypothesis predicted otherwise. Bearing in mind that this group was initially expected to attempt to express more complex ideas, due to their increased WTC, and then go through the proceduralisation processes, it seems plausible to suggest that perhaps this limited intervention did not allow them enough time to go through the latter stages. Perhaps if they were given more time they would show signs of fluency development at a descriptive level as well. Nevertheless, it is important to stress that these observations are only made at a descriptive level and that they cannot provide an answer to the research question.

Looking at the inferential statistics, the only variable offering significant results that contribute to a response of this research question was MLR. The results of MLR showed that TR+ leads to more fluency than no task repetition. This is considered to be an important finding as MLR was characterised as one of the best predictors of fluency in the past (Towell

et al., 1996; Kormos, 2006). Although this result indicates fluency development, the part of the hypothesis predicting fluency development cannot be fully supported. This is because no significant results were found within the experimental conditions. So the findings do not suggest that fluency development came from the incorporation of codeswitching, but rather that it was the result of TR+.

These results cannot be compared to previous empirical studies, as no-one has attempted to investigate the effects of TR+ on fluency. Nevertheless, before interpreting the significant difference of MLR between TR+, which was represented by the two experimental groups, and no task repetition, which was represented by the comparison group, the fact that the two differed in several ways should not be disregarded. Firstly, as it was discussed in the Methodology chapter and in the Quantitative Results chapter, the comparison group was represented by a primary school that was in a different catchment area to that of the experimental groups. While the two areas were not very far away from each other, this difference could have played a role in the students' performance in the intervention. Secondly, the two types of groups differed in terms of their English teacher. The two experimental groups had the same English teacher, while the comparison group had a different one. While an effort was made to ensure that the codeswitching behaviour of the teachers was similar through recordings and observations, which were analysed in terms of the amount and functions of their L1 use, there is always the possibility of the teacher variable affecting the situation. Lastly, the investigation of four potential confounding variables (i.e. proficiency, productive vocabulary, SES & prior L2 contact) showed that the comparison group had a significantly lower level of SES in relation to the two experimental groups. This difference of SES could have also affected the intervention.

Content was measured with verb ratio, L1 syllables and idea units. Although the absence of significant results did not allow for any conclusions to be made, what can be reported as an

interesting finding is the fact that the group under CS conditions did not end up using significantly more L1 syllables than the other two groups. Many researchers and theorists in the past have warned against the excessive and uncontrollable use of the L1 that the allowance of codeswitching may trigger (e.g. Turnbull, 2001). However, this finding can be used as proof that, at least in the context of the present study, the incorporation of codeswitching does not always lead to uncontrollable and excessive L1 use.

Finally, the effects of the intervention on accuracy were investigated with the percentage of error-free clauses and the percentage of correct verb use. The results of both measures showed that TR+ did not have a detrimental effect on accuracy. Only one study examined the effects of TR+ on accuracy in the past. This was the study discussed in the Literature Review, that was conducted by Baleghizadeh and Derakhshesh (2012) in Iran. In this study, the four participants gave a presentation on a book, transcribed the recordings of their own presentation, corrected their own accuracy errors with the help of teachers' corrections and then gave a second presentation. Although the researchers found that the second presentation was more accurate than the first, the possibility of the students' memorising the correct utterances, rather than having actually corrected themselves, was raised. Nevertheless, it could be said that the results of the Baleghizadeh and Derakhshesh's study (2012), as well as the present study, are similar. A positive effect of TR+ is suggested in both studies, although this effect differs to the extent to which it was found to be beneficial. In the present study, accuracy was just retained with TR+, while in the Baleghizadeh and Derakhshesh's study (2012) accuracy was improved.

Another interesting result relevant to accuracy is the fact that the EO group, who received feedback on accuracy, did not do significantly better than the other two groups in terms of accuracy. This result suggested that feedback on accuracy was not effective. However, this result was not surprising for two reasons: first, the topics of the oral production tests were

different to those of the tasks used during the intervention, so a different range of vocabulary was required for the two occasions; and second, the results of the qualitative analysis of students' interactions in Chapter 5 suggested that accuracy feedback could not be easily implemented in the task. Very few instances of students using the correct versions of their errors were identified on subsequent sessions. Perhaps this kind of feedback needs more time to be taken in by the learners and incorporated in their output. Or perhaps it could mean that by simply providing feedback on accuracy, learners are not able to take it in. Follow-up activities could be required for the incorporation of the feedback.

In response to the research question therefore, it was found that TR+ leads to more fluency, (based on MLR only) without undermining accuracy, in relation to no task repetition. This result seems to build on existing research on task repetition. So far, there is limited research suggesting that task repetition has some positive effects on oral production. However, this evidence is not conclusive enough for establishing the pedagogical value of task repetition. The TR+ package therefore has been proposed in the present study as a more viable methodology in real classroom contexts. Learners and teachers might be more willing to accept repeating tasks, when they have a good reason for doing so. From the teachers' interviews, it seems that teachers currently find tasks to be fun activities that are interesting for students to complete, but that they do not offer any significant pedagogical benefit that cannot be gained otherwise. This reason teachers and learners require to justify task repetition could be the incorporation of feedback in the learners' new oral performance. Apart from being more pedagogically acceptable, the *plus* component seems to be effective. The findings of this research question (i.e. enhanced fluency with no detrimental effect on accuracy) provide some initial evidence that this package is effective. As task-based learning research has not been able to establish the value of task repetition so far, perhaps the focus of this research area should turn towards TR+.

7.6. Discussion on RQ5

The final research question inquired about the role of proficiency level in terms of the benefits of TR+. More specifically, it investigated whether HP students benefitted more from TR+ than LP students. All three variable constructs (i.e. fluency, content and accuracy) were investigated in this question as well.

Having controlled for initial differences between the two groups, the results indicated that HP students benefitted significantly more in terms of all temporal fluency measures, idea units, which were a measure of content, and in terms of accuracy. As no other study has investigated the role of proficiency in TR+, direct comparisons with existing findings cannot be made.

Based on the above results, it was argued that TR+ seems to be a condition that works better with learners who have a certain level of proficiency. An explanation for this result could lie in the feedback component. Learners who have a higher level of proficiency might be in a better position to understand the feedback faster and incorporate it in their output instantly. Lower proficiency students might require more time, more feedback or more follow-up activities, as it was suggested for the previous research question, in order to implement it. Future research thus could replicate or conduct a similar version of this study with older learners who are more likely to be a higher level of proficiency overall. Studies with such a sample are more likely to verify the hypothesis put forward in the present study.

7.7. Chapter summary

This chapter summarised all findings emerging from both the qualitative and the quantitative analysis of the present study. These were discussed in relation to the findings of previous research where appropriate.

The first part of the data referred to teachers' opinions on task implementation in the context of the present study. Overall, teachers seemed to be in favour of the implementation of tasks in their classrooms, just like teachers in Carless' studies in Hong Kong. Teachers in Cyprus thought that tasks had a number of benefits on their students, including practical, pedagogical, cognitive and psychological benefits. This finding came in disagreement with Cantonese teachers, who did not think that tasks had more benefits than any other type of activity. Certain problems regarding task implementation were mentioned however, with time consumption being the one mentioned in both contexts.

Teachers also shared their opinion on task repetition and their response seemed unanimous: task repetition should be done with the same type of task, but with different content; an interpretation which was different to that of Bygate (2001) and other researchers. In their interviews however, no reference was made to specific benefits of task repetition on fluency, accuracy or complexity, as it was often investigated in empirical research.

However, the quantitative analysis of the present study provided some evidence suggesting that same-task repetition with feedback was beneficial. More specifically, it was found that task repetition with CS *plus* led to more WTC than task repetition with EO *plus*. Although significant proficiency differences between the two groups were considered, the fact that the CS group did significantly better than both the other two groups in terms of L2 syllables, suggested that the incorporation of L1 in TR+ could lead to more WTC; this was considered

as initial evidence for the argument in favour of this relationship proposed in the Macaro *et al.* (2012) paper.

Similarly, by investigating the effects of the intervention on fluency, content and accuracy as part of RQ4, more results favouring TR+ were found. These suggested that TR+ led to more fluency (based on MLR only) than no task repetition. This result could not be attributed to the incorporation of L1, as no significant differences were found between the CS and the EO group. The initial differences between the experimental groups and the comparison group were considered. In terms of content, although no significant results were found, the fact that the CS group did not use significantly more L1 syllables than the other two groups was important, as it suggested that the incorporation of L1 does not always lead to uncontrollable L1 use. Finally, TR+ did not have a detrimental effect on accuracy; a result similar to that found in the Baleghizadeh and Derakhshesh's study (2012), in which accuracy was actually improved with task repetition with accuracy feedback.

What should be added here as part of the investigation of the intervention are the results of the final research question. The data analysis for the fifth research question showed that TR+ was more beneficial for higher proficiency students than lower proficiency students, in terms of temporal fluency variables, idea units and accuracy. It was concluded therefore that proficiency had a role to play in the effectiveness of TR+ and that future research investigating this package should perhaps recruit older learners, who are more likely to have a higher proficiency level.

A large part of the data in the present study concerned metalanguage. Teachers were asked about metalanguage and, like the authors of previous studies on learner codeswitching, they did not seem to oppose to their students' use of the L1 for metalanguage. In fact, they implied

that this was a natural thing to happen and reported that using the L2 for metalanguage was highly unlikely for their students.

However, the qualitative analysis conducted for RQ2 showed that using the L2 for metalanguage was not such a far-fetched proposition. This analysis examined whether the content of the students' output changed in any way from session 1 to session 2 after TR+. In terms of task language, creativity was found to have both increased and decreased. In terms of L2 metalanguage however, there were two interesting findings: first, the functions of L1 and L2 metalanguage seemed to be common, just as in DiCamilla and Antón's (2012) study, thus supporting the part of hypothesis suggesting that L2 metalanguage can provide the same cognitive support as L1 metalanguage; second, the CS group was able to experiment with L2 metalanguage due to the support of the feedback they received, while no evolution was noted on the part of the EO group. As no previous study looked at the change of language qualitatively when a task was repeated with feedback, comparisons with existing results were hard to be made. Nevertheless, the results provided evidence that using the L2 for metalanguage was possible with these learners.

Finally, the teachers expressed their general preference for mixed-proficiency pairs because they facilitate scaffolding. The qualitative analysis conducted for RQ2 confirmed that scaffolding did take place between the students in this context. It was found that this help came in two forms: in a direct form, in which one student provided the exact utterance or word their partner should use next; and in an indirect form, in which one student pointed out the errors conducted by their partner and allowed time for them to re-structure their sentences.

CHAPTER 8: Conclusion

8.1. Introduction

This chapter presents the conclusions, drawn from the main findings of the present study. These are presented along with their implications for language learning and teaching. The limitations of the study are then discussed, both from a methodological and an analytical perspective. The chapter ends by making recommendations for future work that may build on the theoretical background, as well as the empirical evidence, of this area of research.

8.2. Conclusions and pedagogical implications

The first conclusion that can be drawn from the findings of the present study concerns the TR+ package. As the quantitative results suggested that it helped enhance fluency without undermining accuracy, it is concluded that TR+ seems to be a package that can have positive effects on oral production. With evidence of positive effects, this conclusion can serve as a response to the teachers' concerns about tasks being time-consuming and therefore less favoured when the curriculum has to be covered. Based on the results of the final research question however, what can be added to this conclusion is that TR+ seems to be more effective with learners of a certain proficiency level. One possible explanation of that could be that HP learners are in a better position to instantly incorporate the feedback they receive and its implementation is what contributes to the development of fluency. While this seems to be a more viable package to be used in real classroom settings than task repetition alone, more research and more evidence on this package are needed, before determining its implications for language teaching and learning.

Looking at the two types of the TR+ package separately, more conclusions can be drawn. Concerning the condition including task repetition with codeswitching plus feedback on metalanguage first, it can be concluded that this condition could lead to more WTC. This is based on the fact that the CS group did significantly better than both the other two groups on the L2 syllables measure after the intervention. Additionally, it was concluded that the incorporation of L1 did not lead to excessive L1 use. What can be reported as an implication of these two findings therefore, is that practitioners should be more tolerant with L1 use, as it enhances WTC without leading to uncontrollable L1 use.

Moreover, another important finding emerging from the qualitative analysis of the interactions under this condition is that feedback on metalanguage seems to be enabling learners to maintain the L2 for metalanguage and experiment with it. It has been concluded therefore that this kind of feedback leads to better quality of content when the task is repeated because the feedback provides a good foundation for them to do so. This result matches the aims of task-based learning, as the use and experimentation of L2 metalanguage could be seen as a type of effective L2 communication. Implications for pedagogy based on this conclusion should refer to the perceptions of the teachers, who at least in this context, did not believe that their students were able to use the L2 for metalanguage. Teachers therefore should believe more in the abilities of their students and try to challenge them by introducing this recycling process of L1 metalanguage into the L2 for learners to use as a foundation for target language experimentation.

Concerning the second type of TR+ package, that is English-only conditions and task repetition with feedback on accuracy, conclusions from the data coming from the EO group are drawn. Based on the fact that the EO group did not do better than the other two groups in terms of accuracy, it was concluded that feedback on accuracy is not an effective way to increase accuracy. This was also evident in the qualitative analysis, which suggested that this

kind of feedback was hardly incorporated when tasks were repeated. However, the fact that the extent of giving feedback was limited mainly due to time constraints should be mentioned. If more time was available, then students would be able to receive more feedback and have more time to incorporate it. Nevertheless, what can be reported as a pedagogical implication is that accuracy perhaps needs more input and more practice in order to be incorporated in activities. Follow-up activities may be more effective with accuracy development.

8.3. Limitations

As discussed in the Methodology chapter, there are certain limitations to the present study that are mainly relevant to its methodological aspect. To begin with, the absence of randomisation increased the possibilities of extraneous variables affecting the situation. Although an attempt was made to control for potential confounding variables and match the participants based on these variables in order to make them more comparable, the possibility of other potential extraneous variables being involved cannot be eliminated. Nevertheless, the positive side of using intact classes is that it preserved ecological validity.

Secondly, a potential confounding variable could be the teacher variable, as two different teachers were involved in the intervention. Although an effort was made to control for this variable through observations and recordings that were analysed in terms of the amount and functions of L1 use, their compatibility could still be questioned as not all of their lessons were recorded.

Thirdly, another confounding variable could have been my presence in the intervention classrooms. Although the participants got to know me before the intervention actually started,

the presence of a new person, who is normally not there during their lessons, might have affected their behaviour and their performance during tasks. This was an unavoidable limitation, however, as my presence was required in order to carry out the intervention.

In addition, the non-probability convenience sampling procedure restricted the extent of generalisability of the results. As the sample matched the criteria of the target population, that is sixth grade students in urban Nicosia, generalisations could only be made across the population of the present study.

Moreover, the relatively small sample size might have compromised the significance of the results of the quantitative analysis. As some of the results showed some tendencies at a descriptive level, but were statistically non-significant, perhaps a larger sample size could have helped for these results to be significant at a statistical level as well.

Furthermore, as three different schools participated in the study, the effect of the school cannot be dismissed. This effect could not be controlled for, as no school performance tables are prepared by the Ministry of Education in Cyprus. However, this does not mean that one school cannot be better than the other, in terms of the quality of teaching, the discipline of the students and so on. The possibility of the schools' performance affecting the intervention therefore should be stressed.

Finally, as the sole researcher of the present study, the researcher's bias is likely to have affected both the execution of the study, as well as the analysis of the data. Having this possible effect in mind from the beginning of the study, I made an attempt to remain objective towards all three groups by following the researcher's protocol, described in the Methodology chapter. This helped me ensure that I followed the same steps on every session of every group during the intervention. However, as the goal of the study was to provide

empirical evidence supporting the hypothesis, the possibility of subconsciously favouring the CS group should be mentioned.

8.4. Contribution to the field

Despite these limitations, the present study makes an original contribution to the field of SLA research. It attempts to contribute to the construction of a framework showing when codeswitching is beneficial for language learning and when it is simply the easy option. Although the results have not been as significant and conclusive as I would have hoped for, the insights provided are essential for future studies.

In terms of the field of learner codeswitching, the study investigated the potential effect of the incorporation of codeswitching on oral fluency for the first time. It established this relationship through a hypothesis that is valid both theoretically and empirically. Although the hypothesis was not verified as a whole, the study provided evidence that support parts of this hypothesis. The quantitative analysis showed that allowing L1 use does lead to more WTC, while the qualitative analysis showed that it is possible for L1 metalanguage to be successfully recycled into the L2; the language recycling idea is also one that is original in the field of SLA research. The hypothesis predicted that increased WTC and recycled metalanguage would lead to extensive L2 practice, which would lead to the necessary proceduralisation processes for fluency development. This latter part of the hypothesis was not supported by the data, but it is suggested that perhaps a larger and longer intervention would have allowed more time for proceduralisation processes to take place and eventually develop fluency.

Concerning the field of task-based learning, the present study has promoted the TR+ package, which has been tested by one study in the past, but it has not been looked at in much depth by

theorists in this field. Since this package has yielded some positive results on oral performance in the present study, it is argued that it should be given much more emphasis in the field of task-based learning and studied in much more depth. As studies on task repetition to date have not been able to establish the value of task repetition, perhaps TR+ is the way forward in this field of research. It is also a valid method in terms of the ecology of the classroom. Simple task repetition or massed repetition, which has been studied so far, is more likely to be rejected by practitioners as well as learners. TR+ however, is more likely to be accepted by these populations, as it justifies the repetition of tasks.

8.5. Recommendations for future work

It is hoped that this study has provided several insights for future research. Future replications of this study with larger samples and longer interventions are believed to be able to establish the relationship between codeswitching and oral fluency more effectively by validating the hypothesis.

In terms of the field of task-based learning, future research should focus on the effects of TR+, rather than simply on task repetition, as it was suggested in the previous section. In addition, studies investigating effects on fluency, accuracy and complexity should employ much more complex measures, or combinations of measures, than the ones previously employed, as these are considered to be complex constructs that cannot be adequately captured with simplistic measures.

Finally, the present study is believed to have provided evidence for the argument supporting that numerous variables constantly interact in language acquisition and affect one another. In the present study, almost all the dependent variables underwent some change, although the

focus was initially just on fluency. SLA research therefore, should investigate aspects of language acquisition as groups of variables. The investigation of isolated variables might fail to capture effects on other related variables of equal importance, and thus fail to establish relationships.

References

Ahmadian, M. J. (2011) The effect of 'massed' task repetitions on complexity, accuracy and fluency: does it transfer to a new task? *The Language Learning Journal,* 39 (3), pp. 269-280.

Ahmadian, M. J. & Tavakoli, M. (2011) The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research,* 15 (1), pp. 35-59.

Alley, D. (2005) A study of Spanish II High School students' discourse during group work. *Foreign Language Annals*, 38 (2), pp. 250-257.

Anderson, J. R. (1983) *The Architecture of Cognition,* Cambridge. MA Harvard University Press.

Andon, N. & Eckerth, J. (2009) Chacun á son gout? Task-based L2 pedagogy from the teacher's point of view. *International Journal of Applied Linguistics,* 19 (3), pp. 286-310.

Antón, M. & DiCamilla, F.J. (1999) Socio-cognitive functions of L1 collaborative interaction in L2 classroom. *The Modern Language Journal,* 83 (2), pp. 233-247.

Arnfast, J. S. & Jorgensen, N. J. (2003) Code-switching as a communication, learning, and social negotiation strategy in first-year learners of Danish. *International Journal of Applied Linguistics,* 13 (1), pp. 23-53.

Atkinson, D. (1987) The mother tongue in the classroom: a neglected resource? *ELT Journal,* 41 (4), pp. 241-247.

Atkinson, D. (1993) Teaching in the target language: A problem in the current orthodoxy. *Language Learning Journal,* 8, pp. 2-5.

Babaii, E. & Ansary, H. (2001) The C-test: a valid operationalization of reduced redundancy principle? *System,* 29 (2), pp. 209-129.

Baleghizadeh, S. & Derakhshesh, A. (2012) The effect of task repetition and noticing on EFL learners' oral output. *International Journal of Instruction,* 5 (1), pp. 141-152.

Benson, E.J. (2001) The neglected early history of codeswitching research in the United States. *Language and Communication,* 21, pp. 23-36.

Berlitz (2012). [online] Available at: <http://www.berlitz.co.uk/about_us/tradition/timeline.html> [Accessed 3 August 2012].

Bonk, W.J. & Ockey, G. J. (2003) A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing,* 20, pp. 89-110.

Bradshaw, J. (1990) Test takers' reaction to a placement test. *Language Testing,* 7 (1), pp.13-30.

Broeder, P., Extra, G. & van Hout, R. (1993) Richness and variety in the developing lexicon. In C. Perdue (ed.), *Adult Language Acquisition, Volume II: The Results.* Cambridge: Cambridge University Press, pp.145-163.

Brooks, F. B. (1992) Spanish III learners talking to one another through a jig saw task. *Canadian Modern Language Review,* 48, pp. 696-717.

Brooks, L. (2009) Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language testing,* 26 (3), pp. 341-366.

Brooks, F. B. & Donato, R. (1994) Vygotskyan approaches to understanding foreign language learner discourse during communicative tasks. *Hispania,* 77, pp. 262-274.

Brooks-Lewis, K.A. (2009) Adult Learners' Perceptions of the Incorporation of their L1 in Foreign Language Teaching and Learning, *Applied Linguistics,* 30(2), pp. 216-235.

Bryman, A. (2008) *Social Research Methods (3rd ed.),* Oxford: Oxford University Press.

Burstall, C. (1975) Factors affecting foreign-language learning: a consideration of some relevant research findings. *Language Teaching and Linguistics Abstracts,* 8, pp.105-125.

Butzkamm, W. (2003) We only learn language once. The role of the mother tongue in FL classrooms: death of dogma. *Language Learning Journal,* 28, pp. 29-39.

Bygate, M. (1996) Effect of task repetition: Appraising the development of second language learners. In J. Willis & D. Willis (eds), *Challenge and Change in Language Teaching*. Oxford: Heinemann, pp. 136–146.

Bygate, M. (2001) Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan & M. Swain (eds.), *Researching pedagogic tasks: Second Language Learning, Teaching and Testing.* London: Longman, pp. 23-48.

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*. Amsterdam: John Benjamins, pp. 37–74.

Cambridge Flyers 2: Students' Book (2nd edition) (2007), Cambridge University Press.

Cambridge Flyers 4: Students' Book (2nd edition) (2007), Cambridge University Press.

Canale, M. & Swain, M. (1980) Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics,* 1, pp. 1-47.

Candlin, C. N. (1987) Towards task-based language learning. In C. N. Candlin and D. F. Murphy (eds.), *Language learning tasks.* Lancaster Practical Papers in English. Lancaster and London: Lacaster University.

Carless, D. (2002) Implementing task-based learning with young learners. *ELT Journal,* 56 (4), pp. 389–396.

Carless, D. (2003) Factors in the implementation of task-based teaching in primary schools, *System,* 31, pp. 485-500.

Carless (2004) Issues in teachers' re-interpretation of a task-based innovation in primary schools. *TESOL quarterly,* 38 (4), pp. 639-662.

Carless, D. (2008) Student use of the mother tongue in the task-based classroom. *ELTJ,* 62 (4), pp. 331–338.

Carless, D. & Gordon, A. (1997) Hong Kong primary teachers' perceptions of the difficulties on implementing task-based language teaching. *Journal of Basic Education,* 7 (1), pp. 139-159.

Centeno-Cortés, B. & Jimémez-Jiménez, A. (2004). Problem solving tasks in a foreign language: the importance of the L1 in private verbal thinking. *International Journal of Applied Linguistics*, 14, pp. 7–35.

Chambers, F. (1991) Promoting use of target language in the classroom. *Language Learning Journal,* 4, pp. 27-31.

Charos, C. (1994) *Personality and individual differences as predictors of second language communication: A causal analysis.* Unpublished Honors Thesis, University of Ottawa, Canada.

Clark, J., Lo, Y.C., Hui, M.F., Kam, M., Carless, D. & Wong, P.M. (1999) *An investigation into the development and implementation of TOC initiative with special reference to professional competencies, professional development and resources: Final report.* Hong Kong Institute of Education, Hong Kong.

Cleary, C. (1988) The C-test in English: left-hand deletions. *RECL Journal,* 19 (1), pp. 26-38.

Cohen, J. (1988) *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic Press.

Cohen, J. (1992) A power primer. *Psychological Bulletin,* 112 (1), pp. 155-159.

Cook, V. (2001) Using the first language in the classroom. *Canadian Modern Language Review,* 57 (3), pp. 402-423.

Cook, G. (2010) *Translation in Language Teaching: An argument for reassessment.* Oxford University Press.

Crawford, J. (2004) Language choices in the foreign language classroom: Target language or the learners' first language? *RELC Journal,* 35 (1), pp. 5-20.

Crookes, G. (1989) Planning and interlanguage variability, *Studies in Second Language Acquisition,* 11, pp. 367-383.

De Bot, K. (1992) A bilingual production model: Levelt's 'Speaking' model adapted. *Applied Linguistics,* 13 (1), pp. 1-24.

De la Colina, A. & Mayo, M. (2009) Oral interaction in task-based EFL learning: The use of the L1 as a cognitive tool. *International Review of Applied Linguistics in Language Teaching,* 47 (3-4), pp. 325-345.

DiCamilla, F. J. & Antón, M. (2012) Functions of L1 in the collaborative interaction of beginner and advanced second language learners. *International Journal of Applied Linguistics,* 22 (2), pp. 160-188.

Dörnyei, Z. (2007) *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies.* Oxford: Oxford University Press.

Duff, P.A. & Polio, C.G. (1990) How much foreign language is there in the foreign language classroom? *The Modern Language Journal,* 74 (2), pp. 154-166.

Eckes, T. & Grotjahn, R. (2006) A closer look at the construct validity of C-tests. *Language Testing,* 23 (3), pp. 290-325.

Edstrom, A. (2006) L1 use in the L2 classroom: One teacher's self-evaluation, *The Canadian Modern Language Review,* 63 (2), pp. 275-292.

Eldridge, J. (1996) Code-Switching in a Turkish Secondary School. *ELT Journal,* 50 (4), pp. 303-311.

Ellis, R. (1984) *Classroom second language development.* Oxford: Pergamon.

Ellis, R. (2003) *Task-based language learning and teaching,* Oxford: Oxford University Press.

Ellis, R. (2008) *The Study of Second Language Acquisition.* 2nd ed. Oxford: Oxford University Press.

Ellis, R. (2009) Task-based language teaching: sorting out the misunderstandings, *International Journal of Applied Linguistics,* 19 (3), pp. 221-246.

Field, A. (2009) *Discovering statistics using SPSS* (3rd ed.), London: SAGE Publications.

Foot, M.C. (1999) Relaxing in pairs. *ELT Journal,* 53, pp.36-41.

Foster, P. (1996) Doing the task better: How planning time influences students' performance. In J. Willis and D. Willis (eds), *Challenge and Change in Language Teaching.* Heinemann, pp. 126–135.

Foster, P. (1998) A classroom perspective on the negotiation of meaning. *Applied Linguistics*, 19 (1), pp. 1–23.

Foster, P. & Skehan, P. (1996) The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition,* 18 (3), pp. 299–324.

Foster, P. & Skehan, P. (1999) The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research,* 3, pp. 215-247.

Fotos, S. (2001) Codeswitching by Japan's unrecognised bilinguals: Japanese University students' use of their native language as a learning strategy. *Studies in Japanese Bilingualism,* 22, pp. 329-352.

Freiermuth, M. R. & Huang, H. (2012) Bringing Japan and Taiwan closer electronically: A look at an intercultural online synchronic chat task and its effect on motivation. *Language Teaching Research,* 16 (1), pp. 61-88.

Fulcher, G. (1996) Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing,* 13, pp. 208-238.

Fulcher, G. (2003) *Testing Second Language Speaking,* London: Pearson Longman.

Gass, S., Mackey, A., Fernandez, M. & Alvarez-Torres, M. (1999) The effects of task repetition on linguistic output. *Language Learning,* 49, pp. 549–580.

Gearon, M. (2006) L'alternance codique chez les professeurs de français langue étrangère pendant des leçons orientées vers le développement des connaissances grammaticales. *The Canadian Modern Language Review,* 62 (3), pp. 449-467.

Georgiou, M., Kitromilides, L. & Englezaki, A. (2007) *English for Communication 3: World Surprises (Teacher's book).* Nicosia: Ministry of Education and Culture, Department of Primary Education Curriculum Development Unit.

Ghonsooly, B., Khajavy, G. H. & Asadpour, S. F. (2012) Willingness to communicate in English among Iranian non-English major university students. *Journal of Language and Social Psychology,* 31 (2), pp. 197-211.

Giannikas, C. N. (2011) L1 in English language learning: a research study in a Greek regional context. *International Journal of Applied Linguistics,* 21 (3), pp. 319-339.

Goffman, E. (1974) *Frame analysis: An essay on the organisation of experience.* New York: Harper & Row.

Goldman-Eisler, F. (1951) The measurement of time sequences in conversational behaviour. *British Journal of Psychology,* 42, pp. 355-362.

Goldman-Eisler, F. (1968) *Psycholinguistics experiments in spontaneous speech.* London: Academic Press.

Goldman-Eisler, F. (1972) Pauses, clauses, sentences. *Language and speech,* 15, pp. 103-113.

Gregg, K. (1984) Krashen's Monitor and Occam's Razor. *Applied Linguistics,* 5 (2), pp. 79-100.

Halliwell, S. & Jones., B. (1991) *On target: Teaching in the target language.* London: Centre for Information on Language teaching.

Hancock, M. (1997) Behind Classroom Code Switching: Layering and Language Choice in L2 Learner Interaction. *TESOL Quarterly,* 31 (2), pp. 217-235.

Hashimoto, Y. (2002) Motivation and willingness to communicate as predictor of reported L2 use: The Japanese ESL context. *Second Language Studies,* 20, pp.29-70.

Heinsman, D.T. & Shadish, W.R. (1996) Assignment methods in experimentation: when do nonrandomised experiments approximate answers from randomized experiments? *Psychological Methods,* 1 (2), pp. 154-169.

Hesse-Biber, N. S. & Leavy, P. (2006) *The Practice of Qualitative Research.* California: Sage, Thousand Oaks.

Hilsdon, J. (1991) The group oral exam: Advantages and limitations. In J. C. Alderson & B. North (eds.), *Language testing in the 1990s.* London: Macmillan, pp. 189-197.

Hosoda, Y. (2000) Teacher codeswitching in the EFL classroom. *JALT Journal*, 22 (1), p. 69-93.

Housen, A. & Kuiken, F. (2009) Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics,* 30 (4), pp. 461–473.

Huerta-Macias, A. & Kephart, K. (2009) Reflections on native language use in adult ESL classrooms. *Adult Basic Education and Literacy Journal,* 3 (2), pp. 87-96.

Hymes, D. (1971) *On communicative competence.* Philadelphia: University of Pennsylvania Press.

Hymes, D. (1974) *Foundations in Sociolinguistics: An Ethnographic approach.* Philadelphia: University of Pennsylvania Press.

Issit, S. (2008) Improving scores on the IELTS speaking test. *ELT Journal,* 62 (2), pp. 131-138.

Iwashita, N. (2001) The effect of learner proficiency on interactional moves and modified output in nonnative-nonnative interaction in Japanese as a foreign language. *System,* 29, pp. 267-287.

Iwashita, N., Brown, A., McNamara, T. & O'Hagan, S. (2008) Accessed levels of second language speaking proficiency: How distinct? *Applied Linguistics,* pp. 24-49.

Jacobs, G. M., Dufon, P. & Hong, F. C. (1994) L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading,* 17 (1), pp. 19-28.

Jafarpur, A. (1995) Is C-testing superior to cloze? *Language Testing,* 12 (2), pp. 194-216.

Johnson, W. (1944) Studies in language behaviour: I. A program of research. *Psychological Monographs,* 56, pp.1-15.

Kaponen, M. & Riggenback, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (ed.), *Perspectives on fluency.* Ann Arbor: University of Michigan Press, pp. 5-24.

Kessler, G. (2010) Fluency and anxiety in self-access speaking tasks: the influence of environment. *Computer Assisted Language Learning,* 23 (4), pp. 361-375.

Kim, S.H.O. & Elder, C. (2008) Language choices and pedagogic functions in the foreign language classroom: a cross-linguistic functional analysis of teacher talk. *Language teaching research,* 9 (4), pp. 355-380.

Klein-Braley, C. (1997) C-tests in the context of reduced redundancy testing: an appraisal. *Language Testing,* 14 (1), pp. 47-84.

Klein-Braley, C. & Raatz, U. (1984) A survey of research on the C-Test. *Language Testing,* 1 (2), pp.134-146.

Kormos, J. (2006) *Speech Production and Second Language Acquisition,* New York: Routledge.

Krashen, S. (1982) *Principles and Practice in Second Language Acquisition.* Oxford: Pergamon.

Labov, W. (1972) *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

Lantolf, J.P. (2006) Sociocultural theory and second language learning: a state of the art. *Studies in Second Language Acquisition,* 28 (1), pp. 67-109.

Lee, J. H. (2010) *The differential effects of teacher code-switching on the vocabulary acquisition of adult and young EFL learners: A study in the Korean context.* Unpublished D.Phil. thesis: University of Oxford.

Lee-Ellis, S. (2009) The development and validation of a Korean C-Test using Rasch Analysis. *Language Testing,* 26 (2), pp. 245-274.

Leeming, P. (2011) Japanese high school students' use of L1 during pair work. *International Journal of Applied Linguistics,* 21 (3), pp. 360-382.

Leeser, M.J. (2007) Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning,* 52 (2), pp. 229-270.

Lennon, P. (1990) Investigating oral fluency in EFL: A quantitative approach. *Language Learning,* 40 (3), pp. 387-417.

Levelt, W. (1989) *Speaking with intention to articulation.* Cambridge, MA MIT Press.

Levine, G. (2003) Student and instructor beliefs and attitudes about target language use, first language use and anxiety: Report of a questionnaire study. *The Modern Language Journal,* 87 (3), pp. 343-364.

Li, D. (1998) "It's always more difficult than you plan and imagine": Teachers' perceived difficulties in introducing the communicative approach in South Korea. *TESOL Quarterly,* 32 (4), pp. 677-703.

Littlewood, W. (1981) *Communicative Language Teaching: An introduction.* Cambridge: Cambridge University Press.

Liu, J. (2008) L1 use in L2 vocabulary learning: facilitator or barrier. *International Education Studies*, 1 (2), pp. 65-69.

Liu, D., Ahn, G., Baek, K. & Han, N. (2004) South Korean high school English teachers' code switching: Questions and challenges in the drive for maximal use of English in teaching. *TESOL Quarterly,* 38 (4), pp. 605-636.

Long, M. (1985) Input and second language acquisition theory. In S.M. Gass and C.G. Madden (eds.), *Input in second language acquisition.* Rowley, MA: Newbury House, pp. 377-393.

Lynch, T. & Maclean, J. (2000) Exploring the benefits of task repetition and recycling for classroom language learning. *Language Teaching Research*, 4, pp. 221–250.

Lynch, T. & Maclean, J. (2001) Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan and M. Swain (eds), *Researching Pedagogic Tasks. Second Language Learning, Teaching and Testing.* Harlow, UK: Longman, pp. 99–118.

Macaro, E. (1997) *Target language, collaborative learning and autonomy.* Clevedon: Multilingual Matters.

Macaro, E. (2001) Analysing student teachers' codeswitching in foreign language classrooms: Theories and decision making. *The Modern Language Journal,* 85 (4), pp. 531-548.

Macaro, E. (2005) Codeswitching in the L2 classroom: A communication and learning strategy. In E. Lurda (ed.), *Non-native Language Teachers: Perceptions, Challenges, and Contributions to the Profession.* Boston, MA: Springer, pp. 63-84.

Macaro, E. & Erler, L. (2008) Raising the achievement of young-beginner readers of French through strategy instruction, *Applied Linguistics,* 29(1), pp. 90-119.

Macaro, E., Guo, T., Chen, H. & Tian, L. (2009) Can differential processing of L2 vocabulary inform the debate on teacher codeswitching behaviour? The case of Chinese learning of English. In Daller *et. al.* (eds), *Vocabulary studies in first and second language acquisition: The interface between theory and application.* London: Palgrave Macmillan, pp. 125-146.

Macaro, E. & Lee, J. H. (2013, Forthcoming) Learners' views on codeswitching or English-only instructions: Does age make a difference? *TESOL Quarterly.*

Macaro, E., Nakatani, Y., Hayashi, Y. & Khabbazbashi, N. (2012) Exploring the value of bilingual language assistants with Japanese English as a foreign language learners. *The Language Learning Journal,* pp. 1-14.

MacIntyre, P. D. (1994) Variables underlying willingness to communicate: A causal analysis. *Communication Research Reports,* 11, pp. 135-142.

MacIntryre, P. D. (2007) Willingness to communicate in the second language: understanding the decision to speak as a volition process. *The Modern Language Journal,* 91, pp. 564-576.

MacIntyre, P. D., Baker, S. C., Clément, R. & Conrod, S. (2001) Willingness to communicate, social support, and language-learning orientations of immersion students. *Studies in Second Language Acquisition,* 23, pp. 369-388.

MacIntyre, P. D., Clément, R., Dörnyei, Z. & Noels, K. (1998) Conceptualising willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *Modern Language Journal,* 82, pp. 545-562.

MacIntyre, P. D. & Doucette, J. (2010) Willingness to communicate and action control. *System,* 38, pp. 161-171.

MacIntyre, P. D. & Legatto, J. J. (2011) A dynamic system approach to willingness to communicate: Developing an idiodynamic method to capture rapidly changing affect. *Applied Linguistics,* 32 (2), pp. 149-171.

Malvern, D. D. & Richards, B. J. (2002) Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing,* 19 (1), pp. 85-104.

McCroskey, J.C. (1992) Reliability and validity of the willingness to communicate scale. *Communication Quarterly,* 40, pp. 16-25.

McCroskey, J. C. & Baer, J. E. (1985) *Willingness to communicate: The construct and its measurement*. Paper presented at the annual convention of the Speech Communication Association, Denver, Colorado.

McCroskey, J. C. & Richmond, V. P. (1987) Willingness to communicate. I J.C. McCroskey & J. A. Daly (eds), *Personality and interpersonal communication.* Newbury Park: Sage, pp. 129-156.

McLaughlin, B. (1987) *Theories of Second-Language Learning.* Baltimore: Edward Arnold.

McNamara, T. F. (1997) 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics,* 18(4), pp. 446-465.

Meara, P. & Fitzpatrick, T. (2000) Lex30: an improved method of assessing productive vocabulary in an L2. *System,* 28 (1), pp.19-30.

Meng, Q. (2005) *The Effects of Teacher Codeswitching on Vocabulary Learning in the EFL Classroom*. Unpublished M.Sc. dissertation: University of Oxford.

Milton, J. (2009) *Measuring Second Language Acquisition,* Bristol: Multilingual Matters.

Ministry of Education and Culture: New Curricula, 2011. [online] Available at: <http://www.moec.gov.cy/analytika_programmata/nea-analytika-programmata/anglika_dimotiko.pdf> [Accessed 26 September 2011].

Moore, P.J. (in press) An emergent perspective on the use of the first language in the EFL classroom. *The Modern Language Journal*.

Myers-Scotton, C. (1997) *Duelling Languages: Grammatical structure in codeswitching.* Oxford: Clarendon Press.

Nation, I.S.P. (2001) *Learning Vocabulary in Another Language,* Cambridge: Cambridge University Press.

Norton, J. (2005) The paired formal in the Cambridge Speaking Tests. *ELT Journal,* 59 (4), pp. 287-297.

Nunan, D. (2004) *Task-based Language Teaching*. Cambridge University Press.

Ohta, A. (2001) *Second language acquisition processes in the classroom*: *Learning Japanese.* Mahwah, N. J.: Lawrence Erlbaum.

O'Sullivan, B. (2002) Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing,* 19 (3), pp. 277-295.

Polio, C.G. & Duff, P. A. (1994) Teachers' language use in university foreign language classrooms: a qualitative analysis of English and target language alternation. *The Modern Language Journal,* 78 (3), pp. 313-326.

Poplack, S. (1980) "Sometimes I start a sentence in English y termino en espanol". *Linguistics,* 18, pp. 581-616.

Prabhu, N.S. (1987) *Second Language Pedagogy.* Oxford: Oxford University Press.

Pulido, D. C. (2004) The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, 54 (3), 469–523.

Qian, X., Tian, G. & Wang, Q. (2009), Codeswitching in the primary EFL classroom in China – Two case studies. *System,* 37, pp. 719-730.

Raschka, C., Secombe, P. & Chi-Ling, H. (2009) Conflicts and tensions in codeswitching in a Taiwanese EFL classroom. *International Journal of Bilingual Education and Bilingualism,* 12 (2), pp. 157-171.

Richards, J. C. (2002) Theories of teaching in language teaching. In J. C. Richards & W. A. Renandya (eds.), *Methodology in language teaching: An anthology of current practice.* Cambridge University Press, Chapter 2.

Richards, B. J. & Malvern, D. D. (1997) *Quantifying lexical diversity in the study of language development,* Reading: Faculty of Education and Community Studies, The University of Reading.

Richards, J. C. & Rodgers, T. S. (2001) *Approaches and methods in language teaching,* 2nd ed. Cambridge: Cambridge University Press.

Riggenbach, H. (1991) Towards an understanding of fluency: A micro-analysis of non-native speaker conversation. *Discourse processes,* 14, pp. 423-441.

Robinson, P. (2005) Cognitive complexity and task sequencing: a review of studies in a Componential Framework for second language task design. *International Review of Applied Linguistics in Language Teaching,* 43, pp. 1-32.

Robinson, P., Cadierno, T. & Shirai, Y. (2009) Time and motion: Measuring the effects of the conceptual demands of tasks on second language speech production. *Applied Linguistics,* 30 (4), pp. 533-554.

Rolin-Ianziti, J. & Browlie, S. (2002) Teacher use of learners' native language in the foreign language classroom. *The Canadian Modern Language Review.* 58 (3), pp. 402-426.

Rolin-Ianziti, J. & Varshney, R. (2008) Students' views regarding the use of the first language: An exploratory study in a tertiary context maximizing target language use. *The Canadian Modern Language Review,* 65 (2), pp. 249-273.

Rommerveit, R. (1985) Language acquisition as increasing linguistic structuring of experience and symbolic behaviour control. In J.V. Wertsch (ed.), *Culture, communication, and cognition.* Cambridge: Cambridge University Press, pp. 183-204.

Samuda, V. & Bygate, M. (2008) *Tasks in Second Language Learning.* Palgrave MacMillan.

Segalowitz, N. (2007) Access fluidity, attention control, and the acquisition of fluency in a second language. *TESOL Quarterly,* 41 (1), pp. 181-186.

Segalowitz, N. (2010) *Cognitive bases of second language fluency.* New York: Routledge.

Shumin, K. (2002) Factors to consider: Developing adult EFL students' speaking abilities. In J. C. Richards & W. A. Renandya (eds.), *Methodology in language teaching: An anthology of current practice.* Cambridge University Press, Chapter 18.

Skehan, P. (1990) The relationship between native and foreign language learning ability: educational and linguistic factors. In H. Dechert (ed.), *Current Trends in European Second Language Acquisition Research.* Clevedon: Multilingual Matters.

Skehan, P. (1996) A framework for the implementation of task-based instruction. *Applied Linguistics,* 17 (1), pp. 38-62.

Skehan, P. (1998) *A Cognitive Approach to Language Learning.* Oxford University Press.

Skehan, P. & Foster, P. (1997) Modifying the task: the effects of surprise, time and planning type on task-based foreign language instruction. *Thames Valley Working Papers in ELT,* 4, pp. 38–50.

Spolsky, B. (1973). What does it mean to know a language; or how do you get somebody to perform his competence? In J. Oller & J. Richards (Eds.), *Focus on the learner* (pp. 164-176). Rowley, MA: Newbury House.

Stansfield, C. W. & Kenyon, D. M. (1992) The development and validation of a simulated oral proficiency interview. *The Modern Language Journal,* 72(2), pp. 129-141.

Storch, N. & Aldosari, A. (2010) Learners' use of first language (Arabic) in pair work in an EFL class. *Language Teaching Research,* 14 (4), pp. 355-375.

Storch, N. & Wigglesworth, G. (2003) Is there a role for the use of the L1 in an L2 setting? *TESOL Quartetly,* 37 (4), pp.760-770.

Swain, M. (1985) Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass and C. G. Madden (eds.), *Input in second language acquisition.* Rowley, MA: Newbury House, pp. 235-254.

Swain, M. & Lapkin, S. (2000) Task-based second language learning: the uses of the first language. *Language Teaching Research,* 4 (3), pp. 251-274.

Tang, J. (2000) An empirical study of the use of the mother tongue in the L2 reading class. *Hong Kong Journal of Applied Linguistics,* 5 (2), pp. 45-59.

Taylor, L. & Wigglesworth, G. (2009) Are two heads better than one? *Language testing,* 26 (3), pp. 325-339.

Temple, L. (2000) Second language learner speech production. *Studia Linguistica,* 54 (2), pp. 288-297.

Tian, L. (2009) *Teacher codeswitching in a communicative EFL context: Measuring the effects on vocabulary learning.* Unpublished D.Phil thesis: University of Oxford.

Tien, C. (2009) Conflict and accommodation in classroom codeswitching in Taiwan, *International Journal of Bilingual Education and Bilingualism,* 12 (2), pp. 173-192.

Towell, R., Hawkins, R. & Bazergui, N. (1996) The development of fluency in advanced learners of French. *Applied Linguistics,* 17 (1), pp. 84-119.

Turnbull, M. (2001) There is a role for the L1 in second and foreign language teaching, but..., *The Canadian Modern Language Review,* 57 (4), pp. 531-540.

Van Moere, A. (2006) Validity evidence in a university group oral test. *Language testing,* 23 (4), pp. 411-440.

Vermeer, A. (2000) Coming to grips with lexical richness in spontaneous speech data. *Language Testing,* 17 (1), pp. 65-83.

Vygotsky, L. (1986) *Thought and Language.* Cambridge, MA: MIT Press.

Webb, S. (2008) Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition,* 30 (1), pp. 79-95.

Wigglesworth, G. (1997) An investigation of planning time and proficiency level on oral test discourse. *Language Testing,* 14 (1), pp. 85-106.

Wong Fillmore, L. (1985) When does teacher talk work as input? In S.M. Gass and C.G. Madden (eds.), *Input in second language acquisition.* Rowley, MA: Newbury House, pp. 17-50.

Yanbin, H. & Linjing, Z. (2002) The pedagogical and social functions of codeswitching. *MLTA Quarterly,* 127, pp. 35-46.

Yashima, T. (2002) Willingness to communicate in a second language: The Japanese EFL context. *Institute of foreign language and education and research,* 86 (1), pp. 54-66.

Yashima, T., Zenuk-Nishide, L. & Shimizu, K. (2004) The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning,* 54 (1), pp. 119-152.

Appendix 1: Week 1 Task

Week 1 Task

Unit 10 – Narrative task

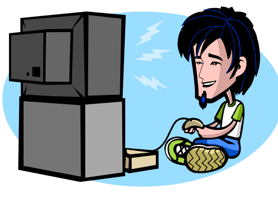
Follow the pictures and tell your partner what you did yesterday. First student A will say what he/she did yesterday and then student B will do the same. Use the verbs above the pictures to help you. Remember! You have to use the verbs in the past tense.

You can start like this.

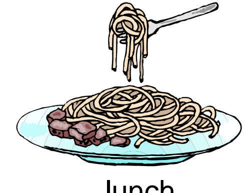
Example: Yesterday, I was at home all day.

Student A

Be – play – have - do – make – have - go



morning



lunch

After lunch

Afternoon

Evening

11:00 p.m.

Unit 10 – Narrative task

Follow the pictures and tell your partner what you did yesterday. First student A will say what he/she did yesterday and then student B will do the same. Use the verbs above the pictures to help you. Remember! You have to use the verbs in the past tense.

You can start like this.

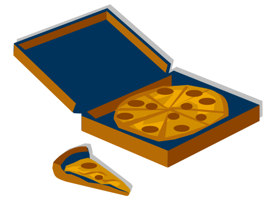
Example: Yesterday, I was not at home.

Student B

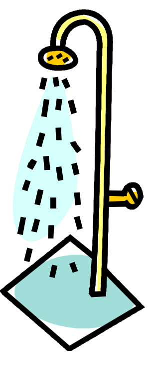
be – go – play – have – come back – eat – have

Afternoon/park

7:00 p.m./home



Dinner



Evening

Appendix 2: Week 2 Task

Week 2 Task

Unit 11 – Information gap task

Make a dialogue with your partner about what you did yesterday. Student A will ask the questions first and Student B will answer. Then, Student B will ask the questions and Student A will answer. Do not look at your partner's sheet!

Example:

Student A

Go/Park

Student B

Go/Mall

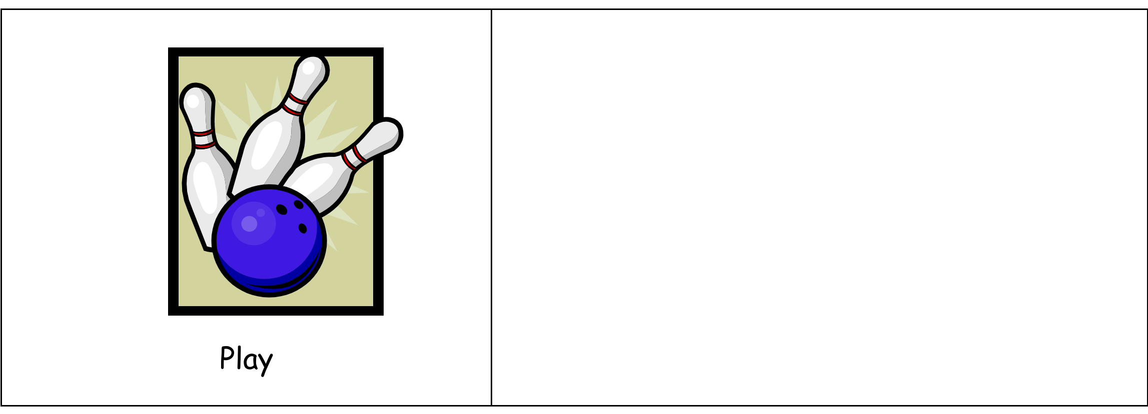
Did you go to the park yesterday?

No, I didn't. I went to the mall yesterday.

Student A

Part A

Student A (Questions)	Student B (Answers)
Go	


Play

Part B - Now you should answer the questions.

Student B (Questions)	**Student A (Answers)**
	Make/cake
	Eat
	homework

Unit 11 – Information gap task

Make a dialogue with your partner about what you did yesterday. Student A will ask the questions first and Student B will answer. Then, Student B will ask the questions and Student A will answer. Do not look at your partner's sheet!

Example:

Student A

Go/Park

Student B

Go/Mall

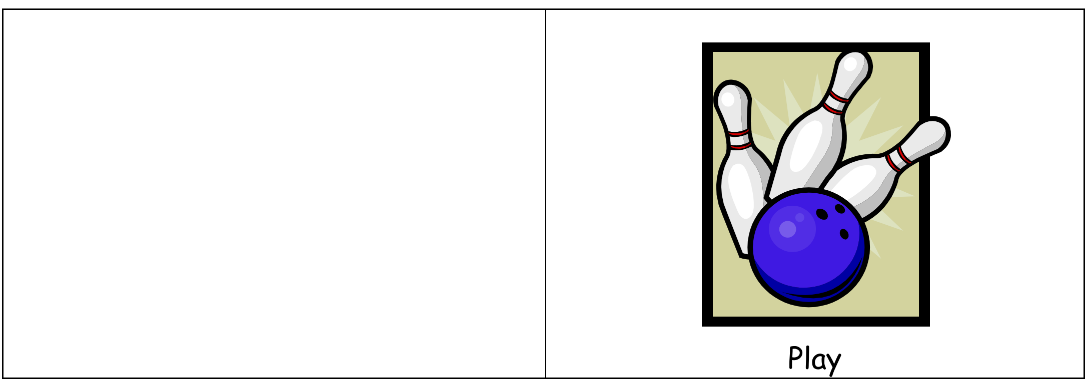
Did you go to the park yesterday?

No, I didn't. I went to the mall yesterday.

Student B

Part A

Student A (Questions)	Student B (Answers)
	computers lesson
	Go



Play

Part B - Now you should ask the questions.

Student B (Questions)	**Student A (Answers)**
Make/biscuits	
Eat	
homework	

Appendix 3: Week 3 Task

Unit 12 – Narrative task

Look at the pictures below and tell the story with your partner. You can use the words below to help you. You don't have to use all the words. But be careful! You have to use the verbs in the past tense. You can use some verbs twice.

What happened to Tom, Sally and Lilly yesterday?

see - poster – go – not have - eat – clown

Cambridge Flyers 2: Students' Book (2nd edition) (2007), Cambridge University Press, p.74.

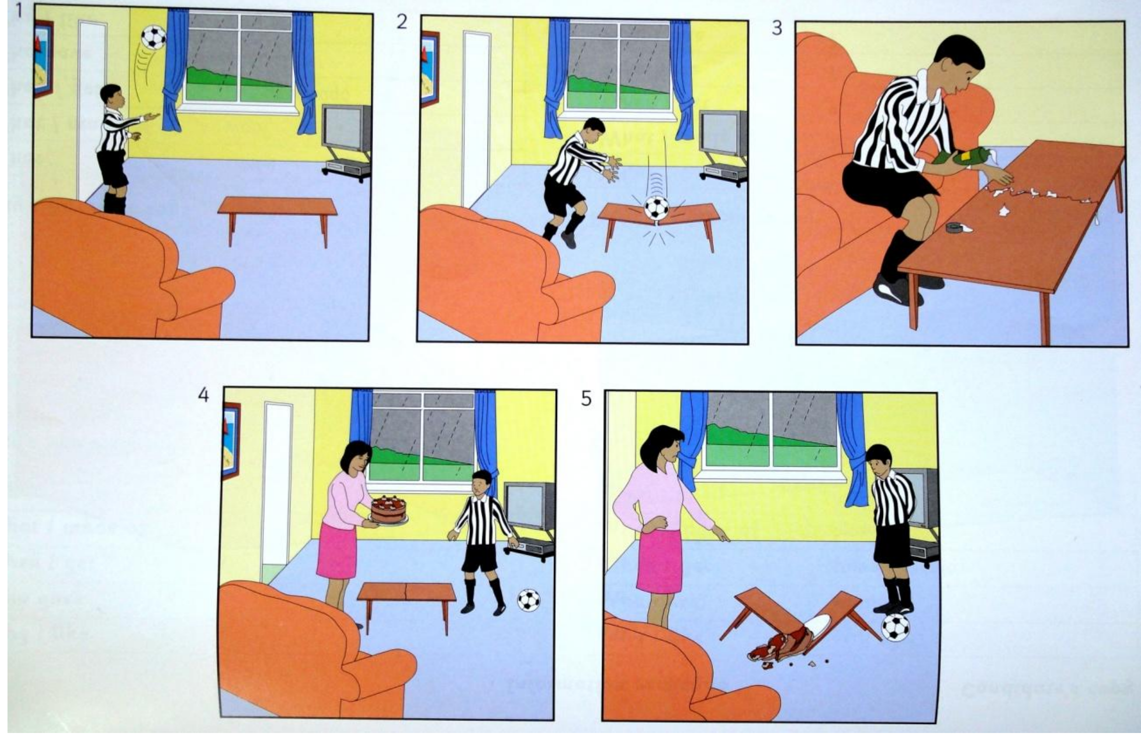
Appendix 4: Week 4 Task

Narrative task

Look at the pictures below and tell the story with your partner. If you want, you can use the words above the pictures to help you. You can narrate the story in your own way. But be careful! You have to use the verbs in the past tense.

What did Billy do yesterday?

fall – break – glue - angry

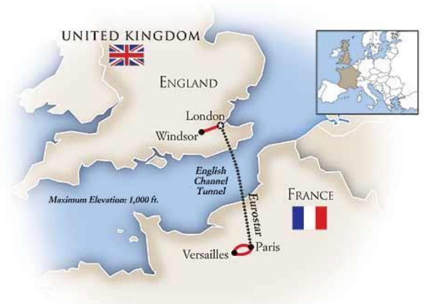


Cambridge Flyers 4: Students' Book (2nd edition) (2007), Cambridge University Press, p. 68.

Appendix 5: Week 5 Task

Unit 13 – Opinion gap task

Imagine that you are going to travel from London to Paris for your Easter holidays. How are you going to travel? Talk with your partner about the means of transport shown in the pictures below. You can use the adjectives below in the comparative and the superlative form. You have to decide how you are going to travel.



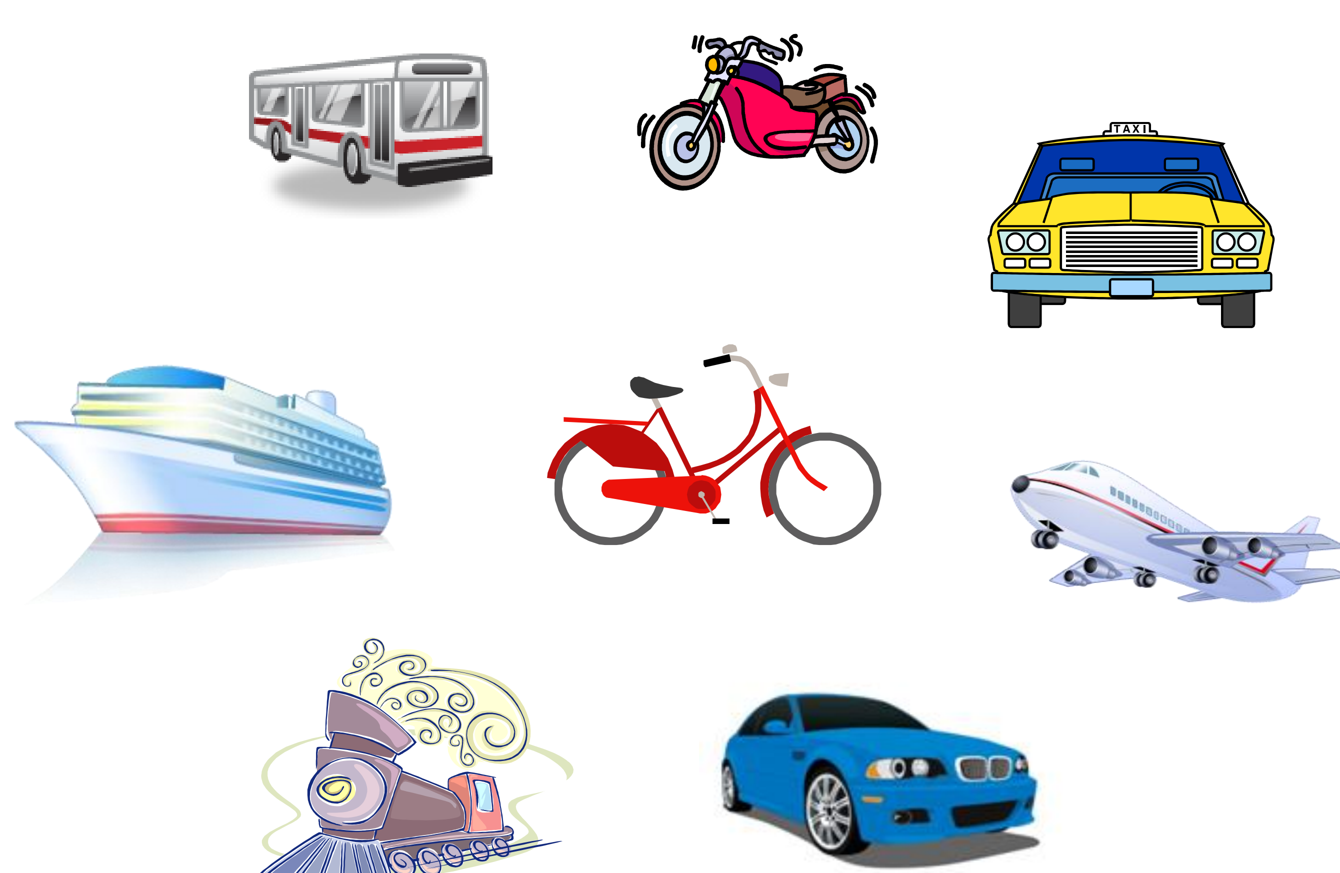
Example:

Student A: I think the best way to go to Paris is by plane.

Student B: No, the plane is the most expensive. Why don't we go by train?

Student A: No, the train is slower. What do you think about the bus?

expensive - dangerous – fast – slow – comfortable – exhausting



Appendix 6: Week 6 Task

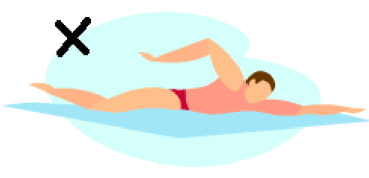
Unit 14 – Doctor/Patient Roleplay

Work with your partner and make a dialogue. Imagine that Student A is sick and goes to the doctor (Student B). The doctor gives some advice to Student A about what he/she should and shouldn't do. Student A asks the doctor what he/she should or shouldn't do. Use the pictures below for ideas, but you can say your own ideas as well!

Example:

Student A: Should I __________?

Student B: No, you shouldn't ________. You should _________.



Appendix 7: Week 7 Task

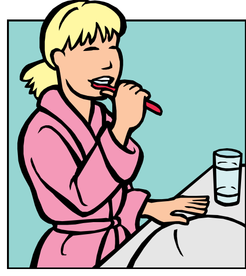
Task 7 – Dentist/Patient Roleplay

Imagine that Student A has a toothache and visits the dentist (Student B). The dentist gives advice to Student A about what he/she should and shouldn't do. Student A asks the dentist what he/she should and shouldn't do. Use the pictures below for ideas, but you can say your own ideas as well!

Example:

Student A: Should I ___________?

Student B: No, you shouldn't _________. You should __________.



Mouthwash

Visit next week

Appendix 8: Week 8 Task

Unit 15 – Information gap task

STUDENT A

Do not show your paper to your partner. You have to try to find each other by giving directions. In Part 1, Student A will give directions to Student B and Student B will try to find Student A. In Part 2, you will do the opposite. Student B will give directions to Student A and Student A will try to find Student B. But remember you have to use the imperative form!

Example:

Student A: Where are you?

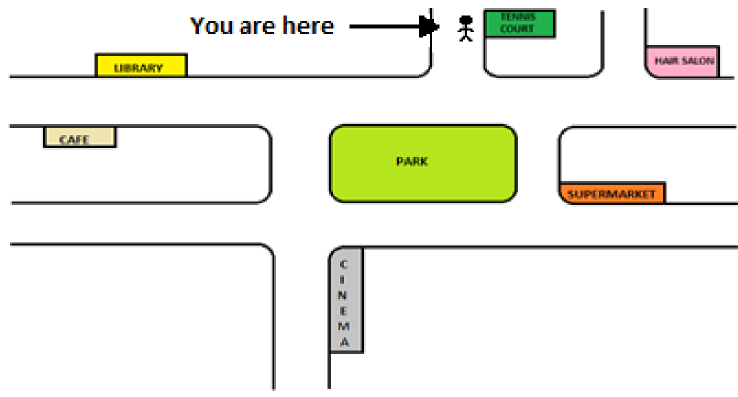
Student B: I am outside the ______________.

Student A: Ok. You have to go straight on. Turn right. Go past the park on your right. Where am I?

Student B: You are at the ______________.

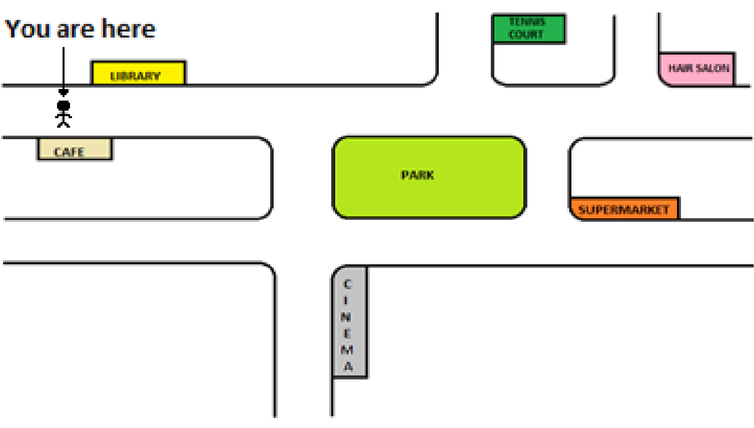
Part 1

You will give directions to Student B to help him/her find you. But first ask him/her where he/she is.



Part 2

Now listen to Student B's directions and try to find him/her. But first tell him/her where you are.



Unit 15 – Information gap task

STUDENT B

Do not show your paper to your partner. You have to try to find each other by giving directions. In Part 1, Student A will give directions to Student B and Student B will try to find Student A. In Part 2, you will do the opposite. Student B will give directions to Student A and Student A will try to find Student B. But remember you have to use the imperative form!

Example:

Student A: Where are you?

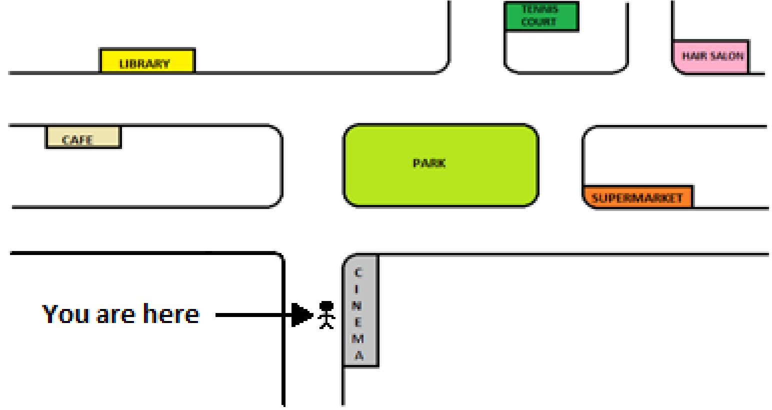
Student B: I am outside the _______________.

Student A: Ok. You have to go straight on. Turn right. Go past the park on your right. Where am I?

Student B: You are at the _______________.

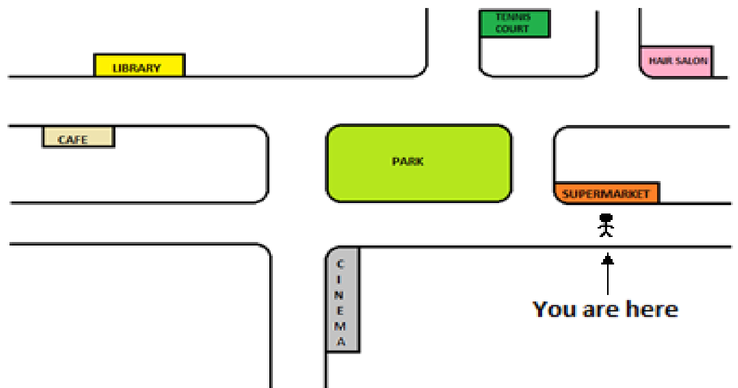
Part 1

Listen to Student A's directions and try to find him/her. But first tell him/her where you are.



Part 2

Now you have to give directions to Student A about how to find you. But first ask him/her where he/she is.



Appendix 9: Fluency handout (Week 5)

What did you say in Greek last time?

We have to make a dialogue.	Πρέπει να κάνουμε διάλογο.
We will compare this with another means of transport.	Θα συγκρίνουμε αυτό με κάποιο άλλο μέσο μεταφοράς.
Let’s end with this picture.	Να καταλήξουμε σε αυτή την εικόνα.
How do we say ___________?	Πως λέμε ________;
What does __________ mean?	Τι σημαίνει ____________;
We already talked about this, this and this picture.	Μιλήσαμε ήδη για αυτή, αυτή και αυτή την εικόνα.
That’s not what we should do. That is the example.	Δεν είναι αυτό που πρέπει να κάνουμε. Αυτό είναι το παράδειγμα.
You have to give your own reasons for each picture.	Πρέπει να πεις τους δικούς σου λόγους για κάθε εικόνα.
We’re done.	Τελειώσαμε.

Appendix 10: Accuracy handout (Week 6)

What is wrong with the following sentences?

*You should stay on the bed.	→ You should stay in bed.
*You don't must go school.	→ You mustn't go to school. You shouldn't go to school.
*Should I have walk with my dog?	→ Should I walk my dog? Should I take my dog for a walk?
*Should I have to wear a jacket?	→ Should I wear my jacket?

Appendix 11: Pre-test

Part A:

I will ask you some questions about your **favourite hobby.**

Part B:

You have one minute to prepare to talk about your **favourite singer or group**. You can take notes to help you remember what you want to say. Try to answer the following questions, but you can add your own ideas about your favourite singer or group:

- Where are they from?
- Are they famous?
- Can you describe them (tall/short, hair, eyes, thin/fat)?
- How old are they?
- Which song do you like best?
- Did you ever go to a concert?
- Anything else you would like to say about them?

Appendix 12: Post-test 1

Part A:

I will ask you some questions about your **favourite film** or your **favourite TV series**.

Part B:

You have one minute to prepare to talk about **your best friend**. You can take notes to help you remember what you want to say. Try to answer the following questions, but you can add your own ideas about your best friend.

- Who is your best friend?
- How old is he/she?
- Where is he/she from?
- Can you describe him/her (hair, eyes, height, weight)?
- When did you first meet?
- Where did you first meet?
- What does he/she like doing?
- How often do you see each other?
- What do you usually do together?
- Why is he/she your best friend?
- Is there anything else you would like to say about your best friend?

Appendix 13: Post-test 2

Part A:

I will ask you some questions about your **best holidays**.

Part B:

You have one minute to prepare to talk about **a person you admire**. This person may be an actor/actress, an athlete, a TV person, a model, a member of your family or anyone else you would like to talk about. You can take notes to help you remember what you want to say. Try to answer the following questions, but you can add your own ideas about this person.

- Who is the person you admire?
- What does this person do?
- Is this person famous?
- How old is he/she?
- Where is he/she from?
- Can you describe him/her (hair, eyes, height, weight)
- When did you first hear about or meet this person?
- Why do you admire this person?
- How do you know about this person? (e.g. TV, magazines, internet)
- Would you like to be like him/her when you grow up?
- Is there anything else you would like to say about this person?

Appendix 14: Excerpts from the transcript of Teacher A's lesson with codeswitching group – Week 4

1. T: I'm going to make a revision on comparisons. Now listen. [...] **Κλείστε τα βιβλία σας *[Close your books].*** Open your small exercise books.
   [Silence]
2. T: I'm going to revise the rules of comparison. Ok? So the first thing. Will you please be quiet? Ok let me just find something. What are these?
3. S: Pencil cases.
4. T: Pencil cases. Very good. This one cost...it cost me more...the purple one cost me 12 euro and the silver one cost me five euros. So we know that when we have two things and we want to compare them what do we say to make comparison? Grace.
5. S: The purple one is cheaper than the purple one.
6. T: So we know that we're gonna use the endings, the –er endings that we already know. So when we've got two things we say the silver one cheaper than the purple. Can I say "this is the cheapest"?
7. S: No.
   [...]

35 S: Can we say this with ink in the pen? This pen has ink. This pen has more ink and this pen has the most ink?

36. T: Yeah you can. **Ναι αλλά όμως *[Yes but]*** you're not using an adjective. We're using adjectives to make...we describe...no we're using **επίθετα *[adjectives]*** adjectives. Ok? We're not using nouns or...we're just talking about how you're gonna use adjectives in comparison....in the comparative form. So this is one case. This is one of the rules.
   [...]

48.T: So when we've got long words that means more than two syllabuses then we have to add "more" or "most". Let's go to...like "cheap" is one syllable. **Μια συλλαβή *[One syllable]***. Ok? "Expensive" it's three. It's a little bit difficult **να κάμεις συλλαβισμό στα Αγγλικά *[to count syllables in English]***. But you do know when it's one syllable. You can realise that. Do you know any words with one syllable? Come on give me some examples.
[...]

60.T: **Να πεις *[To say]*** "most favourite"? **Όπως και το *[like]*** "fun", you can say "more fun". **Αλλά έχει και *[But there is also]*** "funnier".
61.S: "Pretty"
62.T: Pretty. Is it two? Pretty?
63.S: Yes.
64.T: So...**Τώρα να σας πω γιατί έχει ένα άλλο κανόνα δαμέ. *[Let me explain because there is another rule here]* Λόγω του *[Because of the]*** "y". **Τελειώνει σε *[It finishes with a]*** "y".
65.S: [...]
66.T: Ok "good", "bad", "lovely". **Ναι εκείνο πάει *[Yes that is]*** "lovely", "imaginary". **Διαφέρουν...ναι είναι *[They differ yes. It is]*** "love-ly". It's two syllables. It's lovely.
[...]

80. T: Bravo. So it's gonna be "prettier". Ok? And "busy". "Busier". Bow "lovely" it's...it's supposed to be three. No it's one two. Love-ly. So what can you say? Can you say "lovelier", "loveliest"? **Όμως *[But]*** you have to be careful with this because the "y" changes a little bit the word so **εν ακολουθά τους ίδιους κανόνες με τα υπόλοιπα *[it doesn't follow the same rules as the others].*** "More" and "most". Ok? **Οπόταν *[So]*** the rule, the way that you're gonna decide whether you're gonna use "more" and "most" is when you have long words, more than two syllables. Ok? **Γι'αυτό είπα ότι *[That's why I said that]*** some irregular adjectives **κάποια που δεν είναι ομαλά *[some that are not regular]*** they change. Ok? They don't follow the same rules. Is that understood?
81. S: Yes.
82: T: So this is what we will be doing. Now what I want from you is have a little bit talk with your partners and give us an example. Find something on your desk and make a comparison. Your exercise books or your pencil case. You can describe something. Or your hands, or your hair. You can say "my hair is longer than yours" or "I'm more beautiful than you". Just make a decision. Decide.

96. T: Excellent "I am shorter than". Now I would like you to use adjectives that have got more syllables so we can use "more" or "most". So you're gonna give us an example.
97. S: I am thinner than him.
98. T: Yeah but did I just way? What have I just said?
99. S: More syllables.
100. T: **Είπαμε θα χρησιμοποιήσουμε το** ***[We said that we'll use]*** "more" **ή το** ***[or]*** "most".
101. S: His pencil case is more expensive than mine.
102. T: Very good. More expensive. Excellent. Someone else. How about Alex and his partner?
103. S: More handsome.

[...]

110. T: Now "good". **Ποιος να μου πει το** "good" **τι γίνετε?** ***[Who will tell me what "good" becomes?]***
111. S: Better. Best
112. T: Very good. Better, best. And it goes on and on. The other side I think they are all big words. Let's go to the other column. "Popular" what's it going to be? **Πως θα το κάμουμε;** ***[How are we going to convert this?]*** "Popular" **θα γίνει;** ***["Popular" will become?]***
113. S: More popular. Most popular.
114. T: More popular. Most popular. That's it. Now open your books for a while. Leave your exercise books open and let's have a look at exercise 5. [Reads instruction] You see that on each line we've got three items to compare. We have three things to compare. Give an example. Bikes are slower than motorbikes. So you see it compares two things. Ok? But when he had to compare the three of them **δεστε τι έγινε** ***[Look at what it becomes].*** Motorbikes are the most dangerous.

[...]

Appendix 15: Excerpts from the transcript of Teacher B's lesson with comparison group – Week 4

1. T: Last time we started this song remember? You held the paper with the word. When you heard that word you had to raise the paper up in the sky and we said...does anybody remember the title of the song? Does anybody remember the title?
2. S: Harder [low voice].
3. T: **Πέ το *[Say it]*.**
4. S: Harder...
5. T: Right. One of the words was faster. When is it a good thing to be fast? When is it a good thing to be fast? Hm? When you're thinking? Or when you're running in the Olympics? You have to be fast? Who is the fastest man on earth right now? Which athlete is the fastest? Which athlete does this [gesture] when he finishes first? **Εν το πιστεύκω ότι δεν ξέρετε τον Yussein Bald. Ουυυ καλά, καλά.*[I can't believe you don't know who Yussein Bald is. Oh ok].*** Yussein Bald is the fastest man alive right now. Right? Who? Who?
6. S: Επεράσαν τον κυρία *[Miss he's been bitten]*.
7. T: Someone ran faster that Yussein Bald? When did this happen? Recently? Recently? These few past days? Never mind. Now tell me when people mustn't be fast. When is it bad....when is it a bad idea to be fast? When?
8. S: When we do our homework.
9. T: When you do your homework you must be slow? No.
10. S: Να μεν βιαζούμαστε *[We shouldn't rush]*.
11. T: Ok. You have to be careful not to make mistakes. What else? Remember the passage. Remember the kids in the passage. Unit 13. What we studied the previous time. When is it bad to be fast? When you're...
12. S: Driving.
    [....]

17. T: Ok so last time those three kids argued...**διαφωνούσαν *[argued]***...argued whose bike is the best. Right? The oldest, the biggest, the most expensive and they decided to have a race. And what happened? They crashed. They got hurt. Now let's remember some words. Let's remember the comparison. Look at this word: Dangerous. Do you know what it means?
18. S: Επικίνδυνο.
19. T: **Επικίνδυνο *[Dangerous]***. Good. Which one of these do you think that is the most dangerous? You can compare the one with the other and then you can say which one **είναι το πιο πολύ *[is the most]***; the superlative. First you have to compare two things.
20. S: Two?
21. T: **Δύο, ναι *[Two, yes]*.** First compare two. Rock climbing. How do we say this in Greek?
22. S: Ορειβασία *[Rock climbing]*.
23. T: Ok. Rafting? Rafting is when you row this boat in wild water, rivers, falls. **Εν φουσκωτή βάρκα *[It's an inflatable boat]*.** And bungee jumping. You know bungee jumping. So first compare two and then we move on. Who will make a comparison between two of these things? Patrick.
24. P: Rafting is more dangerous than bungee jumping.
25. T: So Patrick thinks that rafting is more dangerous than bungee jumping. Do you agree?
26. S: No.
27. T: What do you think? So you think that bungee jumping is more dangerous than rafting? Ok. Let me hear some other opinions. Each person is entitled to their opinion. **Ατε *[Come on]*.** Fred.
28. F: Bungee jumping is more dangerous than rafting.
29. T: Ok so you agree with Charlie. Tell me about rock climbing. What do you think?
30. S: Είσαι δημμένος *[You're tied up]*.
31. T: No. No you're not.
32. S: Αν ηππέσεις, έππεσες κυρία *[So if you fall, it's over]*.
33. T: Hm, hm. So?
34. S: Έχει σχοινί *[There's a rope].*
35. T: **Πούντο? Α ναι. *[Where is it? Oh yes].*** You're right. Yes, but it's very stiff. It's very stiff. **Ο καθένας έχει την άποψη του *[Each one has their own opinion]*. Δεν έχει σωστό και λάθος *[There's no right or wrong]*.** There is not right or wrong. Each person has one's opinion. Come on tell me about rock climbing. What do you think? **Παιδιά ούτε γνώμη δεν έχετε; *[You don't even have an opinion?]*** Melissa.

36. M: Rock climbing is more dangerous that rafting.
37. T: Ok. Now tell me. Which one do you think that is the most dangerous? More dangerous. The most dangerous [on board] Andrew.
38. A: The most dangerous is rock climbing.
39. T: So he thinks that rock climbing is the most dangerous. Now I want someone to tell me all of these in one sentence. Hm? **Αυτό είναι πιο επικίνδυνο από αυτό αλλά το πιο επικίνδυνο από όλα είναι αυτό. *[This is more dangerous than this but the most dangerous of all is this.]*** Patrick.
40. P: The bungee jumping is...
41. T: We don't say "the". Bungee jumping...
42. P: Bungee jumping is more dangerous than rafting but the most dangerous is rock climbing.
43. T: Good. Very good. Let's move on to the next one. Where is it? "Exciting". What does exciting mean?
44. S: Συναρπαστικό *[Exciting].*
45. T: **Συναρπαστικό *[Exciting].*** Exciting. Who will read the words? Andrew read the words.
46. A: ??
47. T: And?
48. A: ??
49. T: Hiking. Hiking. How do we say hiking in Greek?
50. All: Πεζοπορία *[Hiking].*
51. T: **Πεζοπορία *[Hiking].*** You walk up in the mountains for kilometres. So. I'm listening. More. The most. Come on.

    [….]

59. T: Is there anybody who thinks that hiking is more exciting than fishing? Who thinks that hiking is more exciting than fishing? I do. I don't like fishing. I think it's boring. So for me, for me…for me hiking is more boring than scuba diving, but the most boring of all is fishing. I don't like it. **Γι'αυτό και είπαμε ο καθένας έχει τη γνώμη του *[That's why we said that each one has their own opinion].*** I think you will like this one. Successful. What does it mean? Scott what does successful mean?
60. S: Επιτυχημένος *[Successful].*
61. T: **Επιτυχημένος *[Successful].*** So tell me about who's most successful? Messi, Christiano Ronaldo, Totti? He used to play in Roma. I don't know if he plays there anymore.
62. S: Παιζει κυρία *[He still plays there Miss].*
63. T: **En Roma ακόμα?** *[Is he still at Roma?]* Guys. Wayne.
64. W: Christiano Ronaldo is more successful than Totti.
65. T: But?
66. F: But the...the most successful is Messi.

    [....]

87. T: Laura maybe you think that Greek is more difficult....ee...is easier because it's your language. Your mother tongue, **η μητρική σου γλώσσα *[your mothe tongue].*** Your mother tongue. But if someone else let's say a person from Italy has to learn Greek and English he will think that English is easier than Greek. Or Greek is more difficult than English. Is there another one? This one. I'm sure about the answer. "Interesting". I'm sure all of you will choose reading a book. So Charlie please read it. Hush, hush, silence please.
88. C: Τι σημαίνει "interesting"? *[What does "interesting" mean?*
89. T: So. Interesting. What does "interesting" mean? Natalie what does "interesting" mean? What does it mean? Interesting. There are people who don't know what it means.
90. N: Ενδιαφέρον *[Interesting]*
91. T: **Ενδιαφέρον. Ενδιαφέρον**. So, I'm ready to hear the sentences.

Appendix 16: School letter

[Translated from Greek]

University of Oxford
Department of Education
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024, Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Research student: Maria Vrikki
e-mail: maria.vrikki@education.ox.ac.uk
Tel: 0044xxxxxxxxxxx (UK)/00357xxxxxxxxxxx (CY)

Dear xxxxx,

I am writing to enquire about conducting some research in your school this academic year. I am a doctoral research student at the University of Oxford, supervised by Professor Ernesto Macaro (Professor of Applied Linguistics, Second Language Acquisition), who can be contacted at any time for further information about the project (email: ernesto.macaro@education.ox.ac.uk, tel: 0044 1865 274024). In my research project, *Investigating the impact of different approaches to tasks on L2 oral fluency,* I intend to explore the effects of using the first language during interactive tasks in English lessons.

The research will take place with Grade 6 classes of English during their normal classroom time and will last for approximately four months (February to May 2011). The students participating will be asked to complete a series of tasks in pairs or groups throughout this period. These tasks will be designed as complementary activities based on the units of the book that is currently being used in sixth grade, *English for Communication 3.* One task will be designed for each unit and this will either be a role play, an information gap task or other types of task which are already familiar to the students.

[For Codeswitching group only]: The lessons in the classes that participate will be conducted as normal, up to the point of task completion, which will take place at the end of the lessons. This task will be relevant to what the students will be taught that day. When completing the task, students will be allowed to use Greek, whenever they feel that they cannot express themselves in English. Having collected the phrases students used in Greek, I will introduce the phrases in English during the second lesson of the week. At the end of this lesson, the students will repeat the task and will be encouraged to use the newly introduced language in English. It is believed that this innovative method, which incorporates the first language, could prove to be beneficial in terms of oral fluency.

[For English-only group only]: The lessons in the classes that participate will be conducted as normal, up to the point of task completion, which will take place at the end of the lessons. The students will be asked to complete a task, which will be relevant to what they will be taught that day, by using English only. Then, based on the recordings, I will identify errors made by the students. At the end of the second lesson of the week, the correct versions of these phrases will be presented on a handout and the class will repeat the same task.

[For Comparison group only]: The lessons in the classes that participate will be conducted as normal, up to the point of task completion, which will take place at the end of the first lesson of each week. This task will be relevant to what the students will be taught that day.

This procedure will be repeated for about ten weeks. All interactions between the students when completing tasks, as well as some whole lessons, will be audio-recorded. Also, I will be observing all English lessons during this period. It should be mentioned that by no means

will I be making any judgements in terms of teaching or any changes on what or how the teacher chooses to teach.
Furthermore, before the students start completing tasks, they will be asked to complete in three tests: an oral fluency test, a general proficiency test and a productive vocabulary test. When the students complete all tasks at the end of the ten weeks, they will be asked to complete the oral fluency test again. This test will also be completed for a third time two weeks later. Finally, teachers and students will be interviewed on their experience of participating in the study.

The University of Oxford has strict ethical procedures on conducting ethical research with teachers and young students, consistent with current British Educational Research Association guidelines. This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee, as well as the Ministry of Education and Culture in Cyprus (see enclosed official permission). Before conducting the research, the students' parents or guardians will be sent information pamphlets, which will explain what the study is about and what their children will be asked to do exactly (please see enclosed copy of pamphlet). Parents who agree for their children to participate will be asked to fill in a consent form and return it to the school. In case some parents refuse their children's participation, these children can still be in the classroom, complete the tasks, without being recorded. Having obtained the parents' consent, students will be asked for their own verbal consent. Throughout the research, all participants will be able to withdraw from the study, if they wish to, at any time, without any consequences.

By participating in the research, your school will be contributing to a project that will deepen our understanding of the role of first language use in the foreign language classroom, which is one of the most controversial issues in language acquisition research. Moreover, all students participating will receive a diploma of participation as a reward. In addition, the school will receive a report with the findings of the study.

All participants will remain anonymous in all research reports. The data collected will be kept strictly confidential, available only to my supervisor and myself, and not used other than specified without the further consent of all involved being obtained. All recordings will be deleted at the end of the research period, and kept in secure conditions until then.

If you feel that you would like more information about the study, please do not hesitate to contact me, by using the details at the top of this letter. If you wish to participate, please complete the consent form attached to this letter.

Thank you for your time and attention. I look forward to hearing from you.

Yours sincerely,
Maria Vrikki

Appendix 17: School's consent form

[Translated from Greek]

University of Oxford
Department of Education

Research student: Maria Vrikki
e-mail: maria.vrikki@education.ox.ac.uk
Tel: 0044xxxxxxxxxxxxx (UK)/00357xxxxxxxxxx (CY)

Investigating the impact of different approaches to tasks on L2 oral fluency

Consent form

The purpose of this study is to investigate the effects of different approaches to tasks on oral interaction within the foreign language classroom. The project is undertaken by Maria Vrikki, a doctoral student at the Department of Education, University of Oxford.

1. I have read and understood all the information about this study. I have had the opportunity to ask questions and receive satisfactory answers to my questions.
2. I understand that the school can withdraw from the study without consequence at any time simply by informing the researcher.
3. I understand who will have access to identifying information provided and what will happen to the data at the end of the project.
4. I am aware of who to contact in case I have questions following the school's participation in this study.
5. I understand that this project has been reviewed by and received ethical clearance through the University of Oxford Central University Research Ethics Committee, as well as official permission from the Ministry of Education and Culture in Cyprus.

I agree for the school to participate in this study.

Name: ________________________	Researcher: ____________________
Date: __________________________	Date: _________________________
Signature: ______________________	Signature:______________________

Appendix 18: Teacher's letter

[Translated from Greek]

University of Oxford
Department of Education
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024, Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Research student: Maria Vrikki
e-mail: maria.vrikki@education.ox.ac.uk
Tel: 0044xxxxxxxxxxxx (UK)/00357xxxxxxxxxxx (CY)

Dear xxxxx,

I am writing to invite you and your class of English to participate in my research study, which will be conducted at your school this academic year. I am a doctoral research student at the University of Oxford, supervised by Professor Ernesto Macaro (Professor of Applied Linguistics, Second Language Acquisition), who can be contacted at any time for further information about the project (email: ernesto.macaro@education.ox.ac.uk, tel: 0044 1865 274024). In my research project, *Investigating the impact of different approaches to tasks on L2 oral fluency.* I intend to explore the effects of using the first language during interactive tasks in English lessons.

The research will take place with Grade 6 classes of English during their normal classroom time and will last for approximately four months (February to May 2011). The students participating will be asked to complete a series of tasks in pairs or groups throughout this period. These tasks will be designed as complementary activities based on the units of the book that is currently being used in sixth grade, *English for Communication 3.* One task will be designed for each unit and this will either be a role play, an information gap task or other types of task which are already familiar to the students.

[For CS group only]: The lessons in the classes that participate will be conducted as normal, up to the point of task completion, which will take place at the end of the lessons. This task will be relevant to what the students will be taught that day. When completing the task, students will be allowed to use Greek, whenever they feel that they cannot express themselves in English. Having collected the phrases students used in Greek, I will introduce the phrases in English during the second lesson of the week. At the end of this lesson, the students will repeat the task and will be encouraged to use the newly introduced language in English. It is believed that this innovative method which incorporates the first language could prove to be beneficial in terms of oral fluency.

*[For English-only group only]:*The lessons in the classes that participate will be conducted as normal, up to the point of task completion, which will take place at the end of the lessons. The students will be asked to complete a task, which will be relevant to what they will be taught that day, by using English only. Then, based on the recordings, I will identify errors made by the students. At the end of the second lesson of the week, the correct versions of these phrases will be presented on a handout and the class will repeat the same task.

[For comparison group only]: The lessons in the classes that participate will be conducted as normal, up to the point of task completion, which will take place at the end of the first lesson of each week. This task will be relevant to what the students will be taught that day.

This procedure will be repeated for about ten weeks. All interactions between the students when completing tasks, as well as some whole lessons, will be audio-recorded. Also, I will be observing all English lessons during this period. It should be mentioned that by no means will I be making any judgements in terms of your teaching methods or any changes on what or how you choose to teach.

Furthermore, before the students start completing tasks, they will be asked to complete in three tests: an oral fluency test, a general proficiency test and a productive vocabulary test. When the students complete all tasks at the end of the ten weeks, they will be asked to complete the oral fluency test again. This test will also be completed for a third time two weeks later. Finally, you and your students will be interviewed on your experience of participating in the study.

The University of Oxford has strict ethical procedures on conducting ethical research with teachers and young students, consistent with current British Educational Research Association guidelines. This project has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee, as well as the Ministry of Education and Culture in Cyprus (see enclosed official permission). Before conducting the research, the students' parents or guardians will be sent information pamphlets, which will explain what the study is about and what their children will be asked to do exactly (please see enclosed copy of pamphlet). Parents who agree for their children to participate will be asked to fill in a consent form and return it to the school. In case some parents refuse their children's participation, these children can still be in the classroom, complete the tasks, without being recorded. Having obtained the parents' consent, students will be asked for their own written consent. Throughout the research, all participants will be able to withdraw from the study, if they wish to, at any time, without any consequences.

By participating in the research, your school will be contributing to a project that will deepen our understanding of the role of first language use in the foreign language classroom, which is one of the most controversial issues featuring in language acquisition research. Moreover, all students participating will receive a diploma of participation as a reward. In addition, the school will receive a report with the findings of the study.

All participants will remain anonymous in all research reports. The data collected will be kept strictly confidential, available only to my supervisor and myself, and not used other than specified without the further consent of all involved being obtained. All recordings will be deleted at the end of the research period, and kept in secure conditions until then.

If you feel that you would like more information about the study, please do not hesitate to contact me. If you wish to participate, please complete the consent form attached to this letter.

Thank you for your time and attention. I look forward to hearing from you.

Yours sincerely,
Maria Vrikki

Appendix 19: Teacher's consent form

[Translated from Greek]

University of Oxford
Department of Education

Research student: Maria Vrikki
e-mail: maria.vrikki@education.ox.ac.uk
Tel: 0044xxxxxxxxxxxx (UK)/00357xxxxxxxxxxx (CY)

Investigating the impact of different approaches to tasks on L2 oral fluency

Consent form

The purpose of this study is to investigate the effects of different approaches to tasks on oral interaction within the foreign language classroom. The project is undertaken by Maria Vrikki, a doctoral student at the Department of Education, University of Oxford.

1. I have read and understood all the information about this study. I have had the opportunity to ask questions and receive satisfactory answers to my questions.

2. I understand that I can withdraw from the study without consequence at any time simply by informing the researcher of my decision.

3. I understand who will have access to identifying information provided and what will happen to the data at the end of the project.

4. I am aware of who to contact in case I have questions following my participation in this study.

5. I understand that this project has been reviewed by and received ethical clearance through the University of Oxford Central University Research Ethics Committee, as well as official permission from the Ministry of Education and Culture in Cyprus.

I agree to participate in this study.

Name: _________________________	Researcher: ______________________
Date: ___________________________	Date: ___________________________
Signature: _______________________	Signature:________________________

Appendix 20: Parents' information pamphlet (Codeswitching group)

[Translated from Greek]

Why should you allow your child to participate?

Your child's participation will contribute greatly to this project . If this innovative method of combining the two languages proves to be beneficial, this could have major implications for teaching and learning foreign languages.

The children participating will not deviate at all from their curriculum. Instead they will have fun completing these complimentary tasks. At the end of the study, the children will be given a diploma of participation in order to thank them for their time.

In addition, the school will receive a report with the results of the study.

What do you do now?
If you would like more information about the study, please do not hesitate to contact me, by using the contact details at the back of this pamphlet. I would be happy to talk with you in more detail.

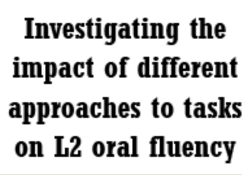
If you wish for your child to participate in the study, please give your written consent by completing the form attached to this pamphlet.

Thank you for you time and help.

Contact details

Maria Vrikki
Tel: 00357 99543348 (CY)
0044 7896312645 (UK)
Email:maria.vrikki@education.ox.ac.uk

Department of Education
15, Norham Gardens
OX2 6PY Oxford
United Kingdom



Investigating the impact of different approaches to tasks on L2 oral fluency

Maria Vrikki

University of Oxford

Department of Education

Dear Parent or Guardian,

My name is Maria Vrikki and I am a doctoral student at the University of Oxford. I am writing to invite your child to take part in my research study, with the rest of your child's class. Your child's school and teacher of English have already agreed to participate in the study. I would appreciate it if you could take a moment to read this pamphlet which informs you about the study.

My research investigates the ways in which the use of Greek in English as Foreign Language classrooms affects interaction, by testing an innovative method which combines the two languages. Bilingual approaches to English learning are becoming increasingly popular across the world.

The study will last for about 10 weeks from January to March 2011.

What will be going on in the classroom?
The English lessons for the classes that participate will run as normal. For the study, the students will be asked to complete a range of different tasks/activities in pairs or groups (e.g. role plays, board games), which will be relevant to what the students will be taught that day. These tasks have been designed based on the book that the school is currently using.

At the end of the first lesson of each week, the students will complete a task and will be allowed to use Greek whenever they feel that they cannot express themselves in English. Then, their teacher, in co-operation with me, will collect the phrases the students used in Greek and introduce them in English during the second lesson of the week. At the end of this lesson, the students will repeat the same task and will be expected to increase their use of English.

A different task will be completed each week and the same procedure will be followed.

During this procedure, I will be present in the classroom and I will be audio-recording the lessons and the conversations of all students completing tasks.

Ethical considerations
The project has been approved by the University of Oxford Central University Research Ethics Committee, as well as the Ministry of Education and Culture in Cyprus.

It is important to mention that all data collected from your child will remain anonymous and will be treated with strict confidentiality. The people who will have access to the data will be myself and my supervisor only.

Please note that your child will have the right to withdraw from the study at any point without any consequences, if she or he wishes to.

Appendix 21: Parents' information pamphlet (English-only group)

[Translated from Greek]

Why should you allow your child to participate?

Your child's participation will contribute greatly to this project . If it is proved that by repeating tasks the oral performance is improved long-term, then major implications could be drawn for teaching and learning foreign languages.

The children participating will not deviate at all from their curriculum. Instead they will have fun completing these complimentary tasks. At the end of the study, the children will be given a diploma of participation in order to thank them for their time.

In addition, the school will receive a report with the results of the study.

What do you do now?

If you would like more information about the study, please do not hesitate to contact me, by using the contact details at the back of this pamphlet. I would be happy to talk with you in more detail.

If you wish for your child to participate in the study, please give your written consent by completing the form attached to this pamphlet.

Thank you for you time and help.

Contact details

Maria Vrikki

Tel: 00357 99543348 (CY)

0044 7896312645 (UK)

Email:maria.vrikki@education.ox.ac.uk

Department of Education

15, Norham Gardens

OX2 6PY Oxford

United Kingdom

Investigating the impact of different approaches to tasks on L2 oral fluency.

Maria Vrikki

University of Oxford

Department of Education

Dear Parent or Guardian,

My name is Maria Vrikki and I am a doctoral student at the University of Oxford. I am writing to invite your child to take part in my research study, with the rest of your child's class. Your child's school and teacher of English have already agreed to participate in the study. I would appreciate it if you could take a moment to read this pamphlet which informs you about the study.

My research investigates the ways in which the use of Greek in English as Foreign Language classrooms affects interaction. Bilingual approaches to English learning are becoming increasingly popular across the world.

The study will last for about 10 weeks from January to March 2011.

What will be going on in the classroom?

The English lessons for the classes that participate will run as normal. For the study, the students will be asked to complete a range of different tasks/activities in pairs or groups (e.g. role plays, board games), which will be relevant to what the students will be taught that day. These tasks have been designed based on the book that the school is currently using.

At the end of the first lesson of each week, the students will complete a task by using English only. The students will be asked to repeat the same task at the end of the second lesson of the week. It has been suggested in research that task repetition is beneficial for language learners in terms of oral proficiency because it allows students to focus on other aspects of language production and improves their oral performance. With this study, I aim to examine these benefits of task repetition.

During this process, I will be present in the classroom and I will be audio-recording the lessons and the conversations of all pairs of students.

Ethical considerations

The project has been approved by the University of Oxford Central University Research Ethics Committee, as well as the Ministry of Education and Culture in Cyprus.

It is important to mention that all data collected from your child will remain anonymous and will be treated with strict confidentiality. The people who will have access to the data will be myself and my supervisor only.

Please note that your child will have the right to withdraw from the study at any point without any consequences, if she or he wishes to.

Appendix 22: Parents' information pamphlet (Comparison group)

[Translated from Greek]

Why should you allow your child to participate?

Your child's participation will contribute greatly to this project . It will help me understand how students approach tasks, which is an integral part of the teaching methodology adopted by language policies in Cyprus, and how they deal with them.

The children participating will not deviate at all from their curriculum. Instead they will have fun completing these complimentary tasks. At the end of the study, the children will be given a diploma of participation in order to thank them for their time.

In addition, the school will receive a report with the results of the study.

What do you do now?
If you would like more information about the study, please do not hesitate to contact me, by using the contact details at the back of this pamphlet. I would be happy to talk with you in more detail.

If you wish for your child to participate in the study, please give your written consent by completing the form attached to this pamphlet.

Thank you for you time and help.

Contact details

Maria Vrikki

Tel: 00357 99543348 (CY)

0044 7896312645 (UK)

Email:maria.vrikki@education.ox.ac.uk

Department of Education
15, Norham Gardens
OX2 6PY Oxford
United Kingdom

Investigating the impact of different approaches to tasks on L2 oral fluency.

Maria Vrikki

University of Oxford

Department of Education

Dear Parent or Guardian,

My name is Maria Vrikki and I am a doctoral student at the University of Oxford. I am writing to invite your child to take part in my research study, with the rest of your child's class. Your child's school and teacher of English have already agreed to participate in the study. I would appreciate it if you could take a moment to read this pamphlet which informs you about the study.

My research investigates the ways in which the use of Greek in English as Foreign Language classrooms affects interaction. Bilingual approaches to English learning are becoming increasingly popular across the world.

The study will last for about 10 weeks from January to March 2011.

What will be going on in the classroom?
The English lessons for the classes that participate will run as normal. For the study, the students will be asked to complete a task/activity at the end of the first lesson of each week. By the end of the study, the students will have completed a range of different tasks in pairs or groups (e.g. role plays, board games). These tasks will be relevant to what the students will be taught that day, since they have been designed based on the book that the school is currently using.

During this procedure, I will be present in the classroom and I will be audio-recording the lessons and the conversations of all students completing the tasks. What I am aiming to do is simply describe the progress that the students make with their oral fluency.

Ethical considerations
The project has been approved by the University of Oxford Central University Research Ethics Committee, as well as the Ministry of Education and Culture in Cyprus.

It is important to mention that all data collected from your child will remain anonymous and will be treated with strict confidentiality. The people who will have access to the data will be myself and my supervisor only.

Please note that your child will have the right to withdraw from the study at any point without any consequences, if she or he wishes to.

Appendix 23: Parents' consent form

[Translated from Greek]

University of Oxford
Department of Education

Research student: Maria Vrikki
e-mail: maria.vrikki@education.ox.ac.uk
Tel: 0044xxxxxxxxxxx (UK)/00357xxxxxxxxxxx (CY)

Investigating the impact of different approaches to tasks on L2 oral fluency

Consent form

The purpose of this study is to investigate the effects of different approaches to tasks on oral interaction within the foreign language classroom. The project is undertaken by Maria Vrikki, a doctoral student at the Department of Education, University of Oxford.

1. I have read and understood all the information about this study. I have had the opportunity to ask questions and receive satisfactory answers to my questions.
2. I understand that my child can withdraw from the study without consequence at any time simply by informing the researcher.
3. I understand who will have access to identifying information provided and what will happen to the data at the end of the project.
4. I am aware of who to contact in case I have questions following my child's participation in this study.
5. I understand that this project has been reviewed by and received ethical clearance through the University of Oxford Central University Research Ethics Committee, as well as official permission from the Ministry of Education and Culture in Cyprus.

I agree for my child to participate in this study.

Name: ____________________________ Researcher: ______________________

Date: _____________________________ Date: ____________________________

Signature: _________________________ Signature: ________________________

Appendix 24: Students' consent slip

[Translated from Greek]

Investigating the impact of different approaches to tasks on L2 oral fluency

Consent form

I have discussed this project with my parents or guardians. They agree for me to participate in the study. I am happy to participate too.

Name: ______________________________ Researcher: ________________________

Date: _______________________________ Date: _______________________________

Signature: ___________________________ Signature: _________________________

Appendix 25: Examples of L1 use under the English-only condition (Pilot study)

The following utterances have been translated in English by the researcher.

METALANGUAGE (OFF-RECORD)		**TASK LANGUAGE (ON-RECORD)**
Metacognitive talk	**Metatalk**	
Define goals: We have to talk about these. **Assign roles:** Who are you? A or B? I will say this. **Decide what to do next:** There is a difference here (Spot the differences talk) Let's say about this. I found something about Wilma Are we done? We already said about this [picture] **Turn taking:** It's your turn. It's my turn now. You should ask now. Continue. **Comment on performance:** We it messed up.	**Identifying sources of difficulty:** How did you pronounce the name? [Pronunciation] How do we say this [in English]? I don't know. What is it? Yes, this is how you should say it [commenting on the partner's language] We are going to use this [structure from the example] You should say "He runs..."	**Unknown vocabulary:** Tail Russia A hundred and twenty Thirty-three

Appendix 26: Examples of L1 use under the Codeswitching condition (Pilot study)

The following utterances have been translated in English by the researcher.

METALANGUAGE (OFF-RECORD)		TASK LANGUAGE (ON-RECORD)
Metacognitive talk	**Metatalk**	
Define goals: [whispering] Here we have to say about [unclear]. So we have to say "No". **Assign roles:** Who are you? A or B? **Decide what to do next:** Do I choose what to say now? Let's move on. Talk about Troodos. There is nothing else to talk about. **Turn taking:** Continue. Go on. There are more pages. It's your turn to ask now. **Comment on performance:** You can. [comments on the partner's performance by saying that s/he is supposed to say "I can"] That's not what you should have said. We messed it up.	**Identifying sources of difficulty:** How do we say "ride a bike"? [asked the researcher] You say: "Why don't we go to...".	**Unknown vocabulary:** Ride a bike Roller-skating Dentist Walk the dog We will fall down Toothache

<u>Appendix 27: C-test</u>

Name: ________________ **Date: ___________________**

Complete the missing parts of the words.

<u>John's family</u>

This is a picture of my family. My fat___ is ta__ and h_'s g__ short bl___ hair.

M__ mother i_ shorter th__ my fat___.

She i_ thin a__ she's g__ long bl___ hair.

I ha__ a sis___ who i_ younger th__ me. She's got blond hair like my mother.

<u>The funny drawing</u>

Nina is a four year-old girl. She ma___ a n__ drawing a__ she sho____ it t__ her par____.

The dra___ is ve__ funny.

T__ elephant i_ smaller th___ the m___.

The butt_____ is big___ than t__ elephant.

T__ girl i__ shorter th__ her do__.

T__ flower i__ taller th___ the tr___. Her parents think that the drawing is beautiful.

<u>In London</u>

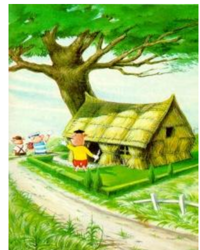
Johnny and his parents are in London. London i_ the cap___ of Eng____ and i_ is o__ of t__ biggest cit___ in t___ world.

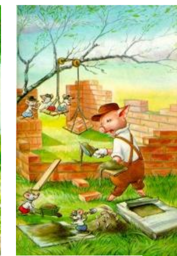
Th___ are ma___ interesting pla____ to vi___ in Lon_____.

Johnny a___ his par____ want t__ visit "Madame Tussaud's".

Th___ is a pl___ where peo____ can s___ wax fig____ of fam____ people.

London is a very interesting place to visit.





The wolf and the three little pigs

Once upon a time there was a big bad wolf and three little pigs.

Each p__ built a ho____. The fi____ house w___ made o_ straw.

T__ second ho____ was ma____ of wo____.

The th____ house w___ made o__ stone.

O__ day t____ big wo____ blew t__ house ma____ of st____ in.

Th____, the wo____ blew t__ house ma____ of wo____ in.

But t__ bad wo____ could n__ blow t__ house ma____ of st___.

So the three pigs lived happily ever after in the third house.

Appendix 28: Productive Vocabulary Test

Name: ______________________ **Date:** _______________________

Class: ______________________ **School:** _____________________

Do you know these words in English?

Ευγενικός *[Polite/Kind]* ____________	Μακρύ *[Long]* ___________
Ενδιαφέρον *[Interesting]* ____________	Ζωολογικός κήπος *[Zoo]*_________
Δασκάλα *[Teacher]* ____________	Διάσημος *[Famous]* ___________
Καινούριος *[New]* ____________	Πανεπιστήμιο *[University]* _______
Φιλικός *[Friendly]* ____________	Πεταλούδα *[Butterfly]* ___________
Ψηλός *[Tall]* ____________	Πρωτεύουσα *[Capital]*___________
Αγενής *[Rude/Impolite]* ____________	Χθες *[Yesterday]* ___________
Έξυπνος *[Clever/Smart]* ____________	Τιμή *[Price]* ___________
Μάτια *[Eyes]* ____________	Εφημερίδα *[Newspaper]*__________
ʻΟμορφη *[Beautiful/Pretty]*___________	Πόλη *[City/Town]* ___________
Ζωγραφιά/Σχέδιο *[Drawing]*__________	Βρίσκω *[Find]* ___________
Γονείς *[Parents]* ____________	Χαρτί *[Paper]* ___________
Βαρύ *[Heavy]* ___________	Χρυσό *[Gold]* ___________
Χιονάνθρωπος *[Snowman]* _________	Πέτρα *[Stone]* ___________
Άτακτος *[Naughty]* ___________	Μάθημα *[Lesson]* ___________

Appendix 29: SES Questionnaire

[Translated from Greek]

If you wish that your child participates in the study, please fill in this short questionnaire which concerns the socio-economic status of the family. The information that you provide will remain confidential and will be used anonymously only for the present study.

Child's name: ___________________________

Part A: Educational level

	Father	Mother
1) Did not attend school at all	☐	☐
2) Primary school certificate	☐	☐
3) Secondary school certificate	☐	☐
4) Tertiary Education School certificate	☐	☐
5) University undergraduate degree	☐	☐
6) Postgraduate degree (Masters or Doctoral)	☐	☐

Part B: Occupational Status

	Father	Mother
1) Public sector employee	☐	☐
2) Private sector employee	☐	☐
3) Self-employed	☐	☐
4) Pensioner	☐	☐
5) Unemployed	☐	☐

Part C: Family income

Net family income per month:

1) Less than €2000 ☐
2) €2000 - €4000 ☐
3) €4000 - €6000 ☐
4) €6000 - €8000 ☐
5) Above €8000 ☐

Thank you for your co-operation.

Appendix 30: Student Questionnaire

[Translated from Greek]

STUDENT QUESTIONNAIRE

Name: ______________________________ School: __________________________

	Yes	No
➢ **Have you ever lived in a country other than Cyprus?**	☐	☐

If yes: Which country? ______________

For how many years? ______________

How old were you when you lived there? ____________

➢ **Is English the mother tongue of one or both your parents?**	☐	☐
If yes, do you speak English at home?	☐	☐

➢ **Have you been taking English lessons outside the school?**	☐	☐

If yes, for how many years? ______________________

➢ **Do you have a contact with the English language in any other way (e.g. English-speaking pre-primary school or summer school/summer camp, friends or relatives or other people who speak English with you)?**

☐ ☐

If yes, please explain: ____________________________________

____

Thank you for your time!

Appendix 31: Researcher’s protocol

<table>
<tr><th colspan="4">PROTOCOL</th></tr>
<tr><th></th><th>CODESWITCHING GROUP</th><th>ENGLISH-ONLY GROUP</th><th>COMPARISON GROUP</th></tr>
<tr><td rowspan="3">LESSON 1</td><td colspan="3">• Give out task and recorders.
• Read out the English instructions once.
• Explain in Greek what they have to do.
• Make sure everyone understands.</td></tr>
<tr><td>Tell them that they can switch to Greek to talk about how to complete the task, when they feel that they cannot express themselves in English.</td><td>Stress that they should only use English when completing the task, even when they talk about how to complete the task.</td><td>(No language instructions)</td></tr>
<tr><td colspan="3">• Give them 10 minutes to complete the task.
• Make sure they identified themselves in the recording.
• When they are done, collect the recorders.</td></tr>
<tr><td rowspan="2">LESSON 2</td><td>• Distribute handout.
• Go through the phrases once. Let the class read out the answers if there is time.</td><td>• Distribute handout
• Go through the errors once
• Make sure they understand the difference between the wrong answer and the correct answer</td><td rowspan="2">(no repeat of task)</td></tr>
<tr><td colspan="2">• Give out the same task.
• Make sure students remember what they have to do.
• Start recorders.
• Encourage students to use the newly introduced language.
• Give them 10 minutes to complete the task.
• Make sure they identified themselves in the recording.
• When they are done, collect the recorders.</td></tr>
</table>

Appendix 32: Teacher letter (Interviews)

[Translated from Greek]

University of Oxford
Department of Education
15 Norham Gardens, Oxford OX2 6PY
Tel: +44(0)1865 274024, Fax: +44(0)1865 274027
general.enquiries@education.ox.ac.uk www.education.ox.ac.uk

Research student: Maria Vrikki
e-mail: maria.vrikki@education.ox.ac.uk
Tel: 0044xxxxxxxxxxxx (UK)/00357xxxxxxxxxxx (CY)

Investigating the impact of different approaches to tasks on L2 oral fluency

Dear xxxxx,

I am a research student in Applied Linguistics at the Department of Education of the University of Oxford and I am writing to invite you to participate in my research study. The study is being supervised by Professor Ernesto Macaro (Professor of Applied Linguistics, Second Language Acquisition), who can be contacted at any time for further information about the project (email: ernesto.macaro@education.ox.ac.uk, tel: 0044 1865 274024).

One of the aims of my study is to explore the opinions of the teachers of English language on the role of tasks in foreign language education. Your participation in the study will involve an interview, which will be audio-recorded. Please note that my purpose is not to judge nor evaluate your way of teaching, but to understand your opinion on this matter.

The University of Oxford has strict ethical procedures on conducting ethical research in schools. The present study has been reviewed by, and received ethical clearance through, the University of Oxford Central University Research Ethics Committee, as well as the Ministry of Education and Culture in Cyprus. Throughout the research, all participants will be able to withdraw from the study, if they wish to, at any time, without any consequences. All collected data will remain anonymous and will be kept strictly confidential. The data will not be used for any other purpose apart from the present study. The only people with access to the data will be me and my supervisor. All recordings will be deleted at the end of the research period, and kept in secure conditions until then.

All participants will in all research reports. The data collected will be kept strictly confidential, available only to me and my supervisor and not used other than specified without the further consent of all involved being obtained. The data collected will be kept strictly confidential, available only to me and my supervisor and not used other than specified without the further consent of all involved being obtained. All recordings will be deleted at the end of the research period, and will be kept in secure conditions until then. It is also worth mentioning that you will have the right to withdraw from the study, if you wish to, at any time, without any consequences.

If you feel that you would like more information about the study, please do not hesitate to contact me. If you wish to participate, please complete the consent form attached to this letter.

Thank you for your time and attention. I look forward to hearing from you.

Yours sincerely,
Maria Vrikki

Appendix 33: Interview Schedule

Task implementation

- Do you usually choose to do the tasks from the book in your classroom?
- Do you think that student-centred activities in general can easily be implemented in Cyprus' school settings?
- Do you think teachers in Cyprus are adequately trained to use tasks in the classroom?
- Do you think that student-centred activities can be beneficial to your students? If yes, in what ways? If not, why not?

Task repetition

- What do you think about task repetition?
- Do you think that students can gain something from task repetition?
- How do you think students would react to repeating activities?

Metalanguage

- Do you think metalanguage is more likely to occur in their mother tongue or in the target language?
- Do you think that metalanguage can be taught in English and still serve the same functions during task completion?

Pairs

- Given that we have mixed-ability classrooms, do you think it is better to have same-ability pairs/groups or mixed-ability pairs/groups?
- *Interlocutor familiarity.* Do you think that letting students choose their partners, rather than you assigning them in pairs, affects the result of the activity?
- *Pupil involvement.* Do you think that there can be a difference in the extent to which students within a pair or group get involved in a task?
- *Scaffolding*. Do you think that it is possible for a more proficient student to provide help to a less proficient student?

Appendix 34: Additional examples of increased and reduced creativity of task language from session 1 to session 2 (Chapter 5)

Transcript 1: Example of increased creativity at S2 from CS group on week 7: Gary (HP) & Peter (LP)

Session 1

1. P: Ποιος είσαι εσύ; *[Who are you?]*
2. G: Huh?
3. P: Εσύ είσαι ο γιατρός; *[Are you the doctor?]*
4. G: Εν με γιατρό πάλε; *[Is it with the doctor again?]*
5. P: Dentist. Hi doctor.
6. G: Hi.
7. P: I have a toothache.
8. G: Why?
9. P: Because I eat...
10. G: "I ate".
11. P: I ate...eee....
12. G: "Many sweets".
13. P: Very...
14. G: "Many sweets".
15. P: Many sweets.
16. G: Why did....Did you brush your teeth? "No".
17. P: No.
18. G: Και μετά ρωτάς *[And then you ask]*: "Should I eat sweets"?
19. P: Should I eat sweets?
20. G: No, you shouldn't. You should eat fruit.
21. P: Should I eat chocolate?
22. G: No, you shouldn't. Only sometimes, if you brush your teeth.
23. P: Should I eat ice-cream?
24. G: No, you shouldn't. And after you brush your teeth, you will have to put mouth....mouthwash in your mouth.
25. P: Should I eat...ate a cake?
26. G: "Should I eat a cake"
27. P: Should I eat cakes?
28. G: <u>We said no sweets. What...what can't you understand of that. Well anyway.</u> You have to visit me next week.
29. P: Goodbye doctor. Thank you.
30. G: Bye.

Session 2

1. G: Πως τον λέμε τον πονόδοντο; *[How do we say "toothache"?]*
2. P: "Toothache".
3. G: I've got a toothache.
4. P: "Well done". Why?
5. G: "<u>Because I told you not to eat sweets</u>."
6. P: Because I told you...
7. G: "Not to eat"
8. P: To not eat.
9. G: "Not to eat"
10. P: Not to eat sweets.

11. G: “Did you eat sweets today?”
12. P: Did you eat sweets today?
13. G: Today, yesterday, two days ago. “You mustn’t eat sweets”
14. P: You mustn’t eat sweets. You should eat fruits.
15. G: I don’t like fruits. Should I eat chocolate?
16. P: No. You should brush your teeth every day.
17. G: Should I eat ice-cream?
18. P: No. You should...
19. G: “Wash your mouth.”
20. P: Wash your mouth.
21. G: “With a mouth wash”.
22. P: With a mouthwash,
23. G: Should I eat cakes? “We said no sweets” πε μου *[Tell me: “We said no sweets”]*.
24. P: We said no sweets. Come to visit next week.
25. G: Ok.
26. P: Bye.

Transcript 2: Example of reduced creativity at S2 from EO group on week 7: Melanie (HP) & Miranda (LP)

Session 1

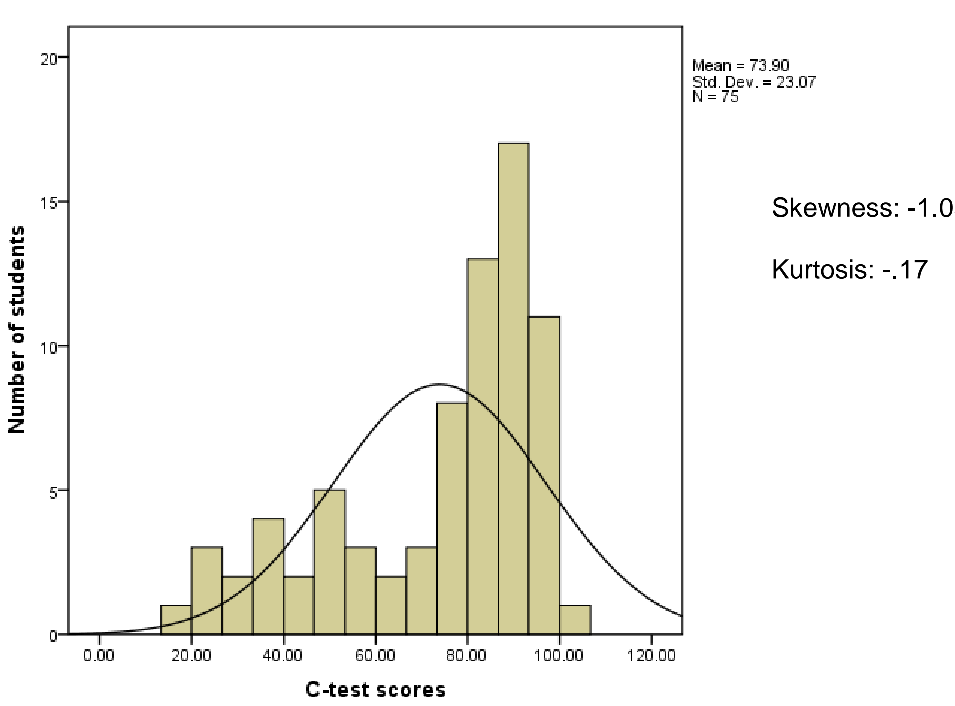
1. Mir: Student A?
2. Mel: You are student A.
3. Mir: No, you are student A.
4. Mel: Should I eat chocolate?
5. Mir: No, you shouldn’t.
6. Mel: “Eat chocolate”.
7. Mir: Eat chocolate. You should eat fruit. Fruits sorry.
8. Mel: Ok. Should I eat ice-cream?
9. Mir: No. That’s not a good idea for someone who...who wants to keep your mouth clean.
10. Mel: Ok then. Should I...ok...should I...should I eat cake? Chocolate cake?
11. Mir: No, you shouldn’t. That’s the reason...that’s the same reason with ice-cream. But you should visit next week the...
12. Mel: “Me”.
13. Mir: Hm?
14. Mel: “Me”. You should come next week...
15. Mir: But you should come next week to visit me.
16. Mel: Ok. Should I have to use mouthwash after dinner?
17. Mir: After dinner, breakfast, lunch. After I eat something, it’s not writing this on the paper but yes you should.
18. Mel: Should I have to...?
19. Mir: Should I have to?
20. Mel: Should I...ok, ok...should I...
21. Mir: “Wash my face?”
22. Mel: Should I, should I wash my teeth after dinner, breakfast and lunch?
23. Mir: Yes. Bye, bye.
24. Mel: What….I will not come to you tomorrow. Next week. I will go to other dentist. [laughs]

Session 2

1. Mel: Should I should I eat candy?
2. Mir: Πούντο candy? A. *[Where is the candy? Oh]* No you shouldn't eat candy. You should eat food.
3. Mel: Should I eat chocolate?
4. Mir: No you shouldn't eat a chocolate. You should use a mouthwash.
5. Mel: Should I use mouthwash?
6. Mir: I said yes.
7. Mel: But no chocolate. Mouthwash.
8. Mir: I said no you shouldn't eat chocolate. You should use mouthwash.
9. Mel: Should I brush my teeth?
10. Mir: Yes you should.
11. Mel: Should I should I eat ice-cream?
12. Mir: No you shouldn't but you should visit next week the dentist.
13. Mel: The dentist. But you are the dentist. So you say: "You have to come next week"
14. Mir: You should come next week. You should visit me next week.
15. Mel: Ok but should I should I eat cake? <u>Because you know I like cake.</u>
16. Mir: No you shouldn't.
17. Mel: Ok bye bye.

Appendix 35: Normality distributions for possible confounding variables data.

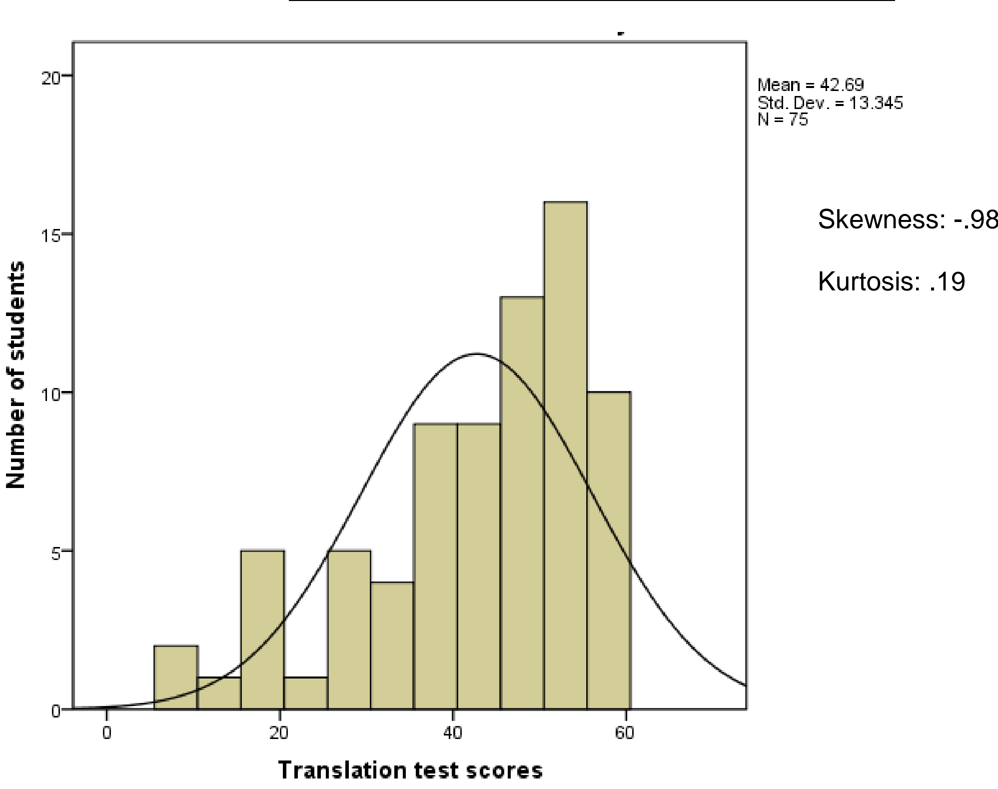
Proficiency scores (C-test)



Skewness: -1.03

Kurtosis: -.17

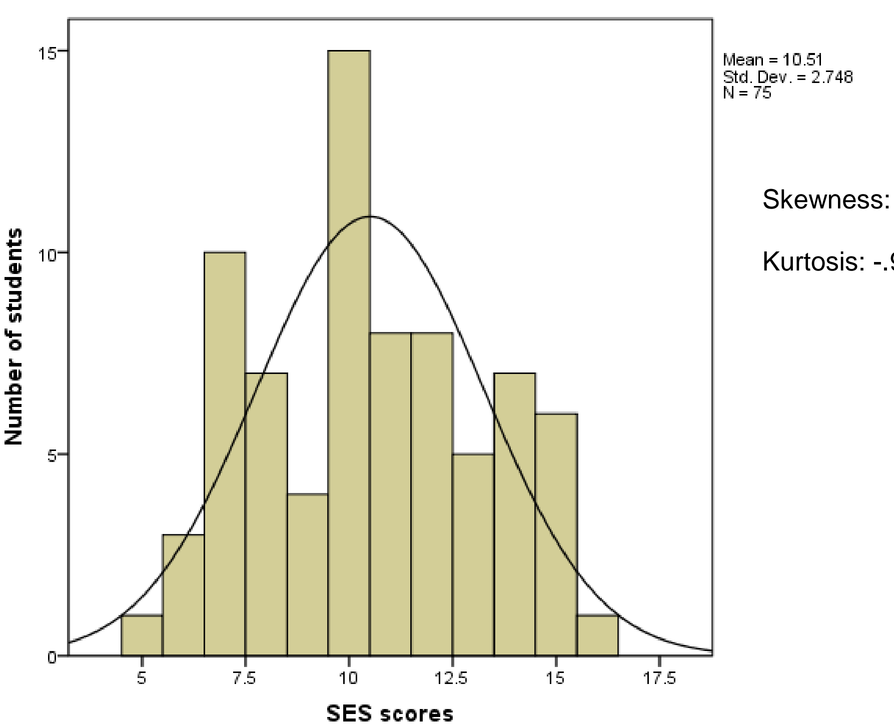
Productive Vocabulary scores (Translation test)



Skewness: -.98

Kurtosis: .19

SES Total score



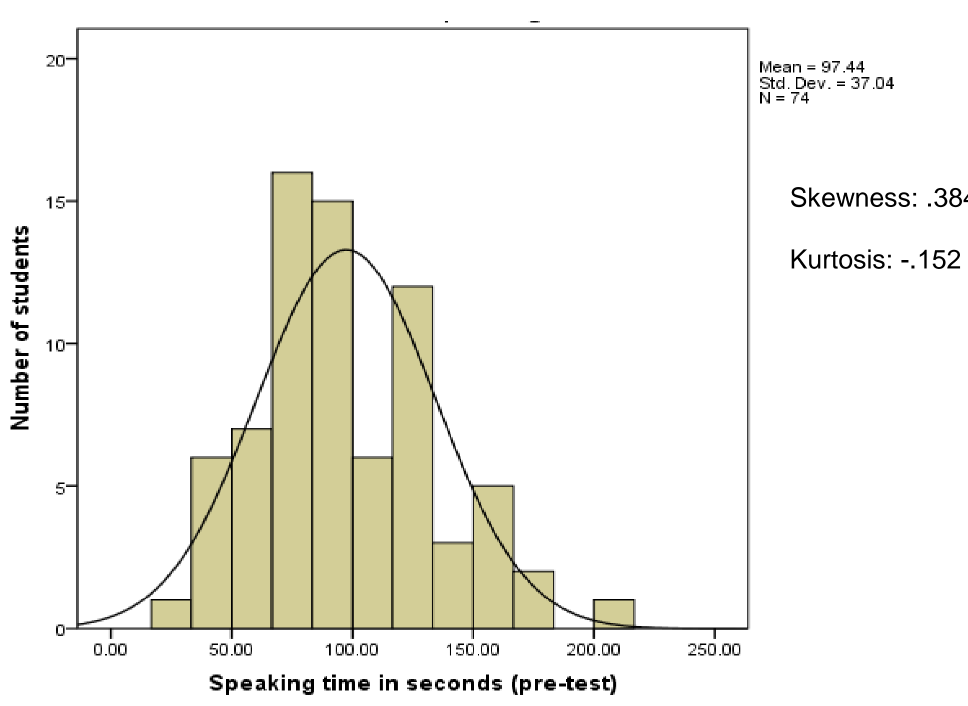
Skewness: .08

Kurtosis: -.91

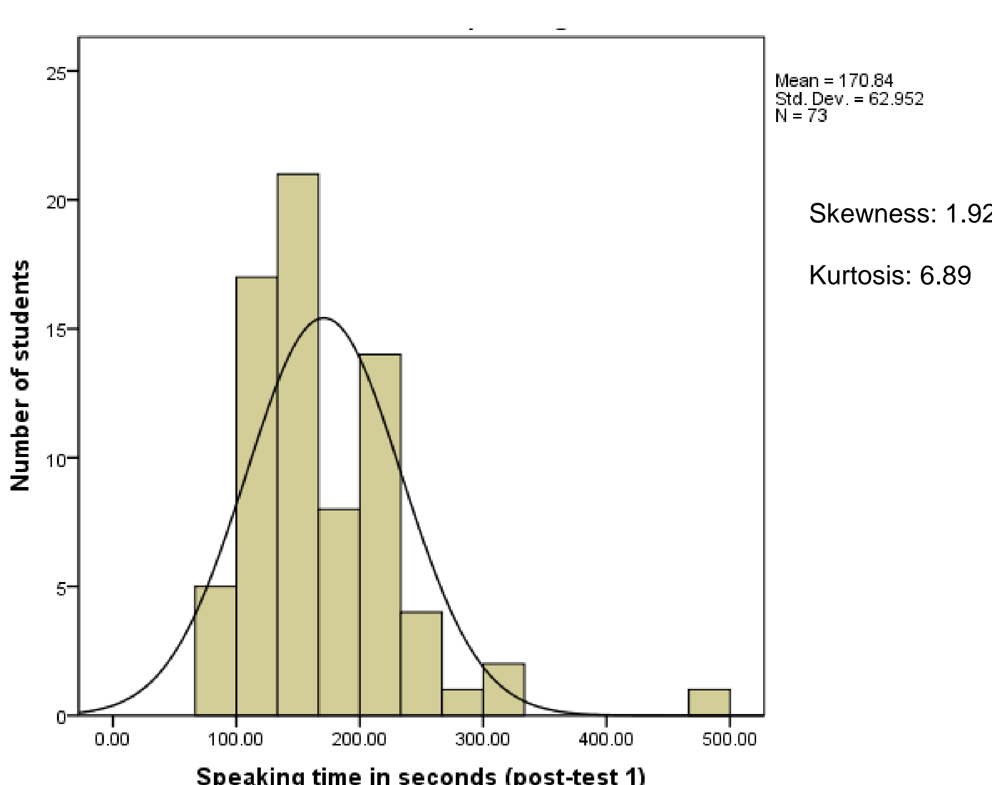
Appendix 36: Distribution graphs for WTC measures

Speaking time

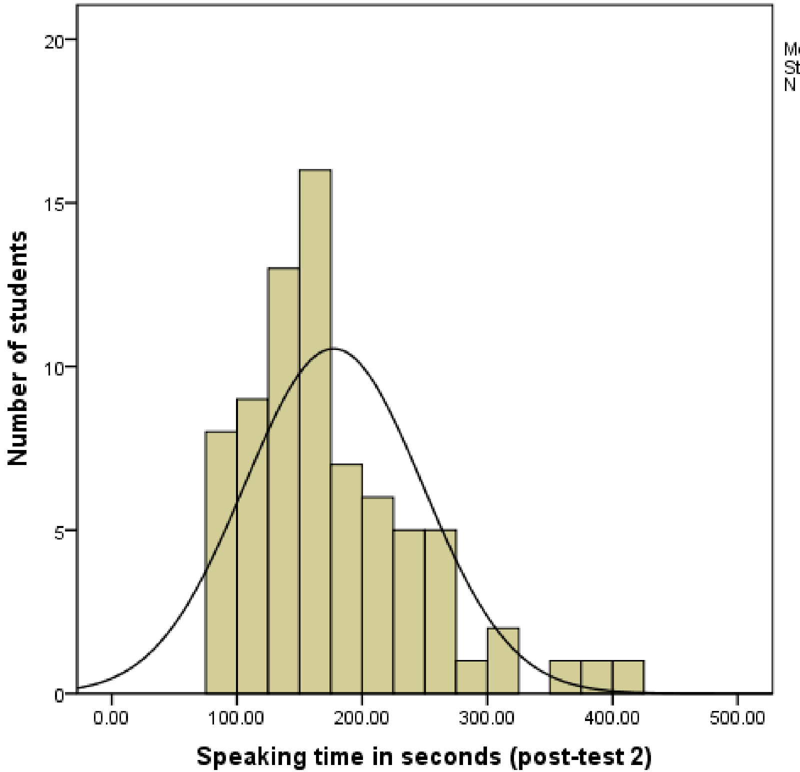
Pre-test



Post-test 1



Post-test 2

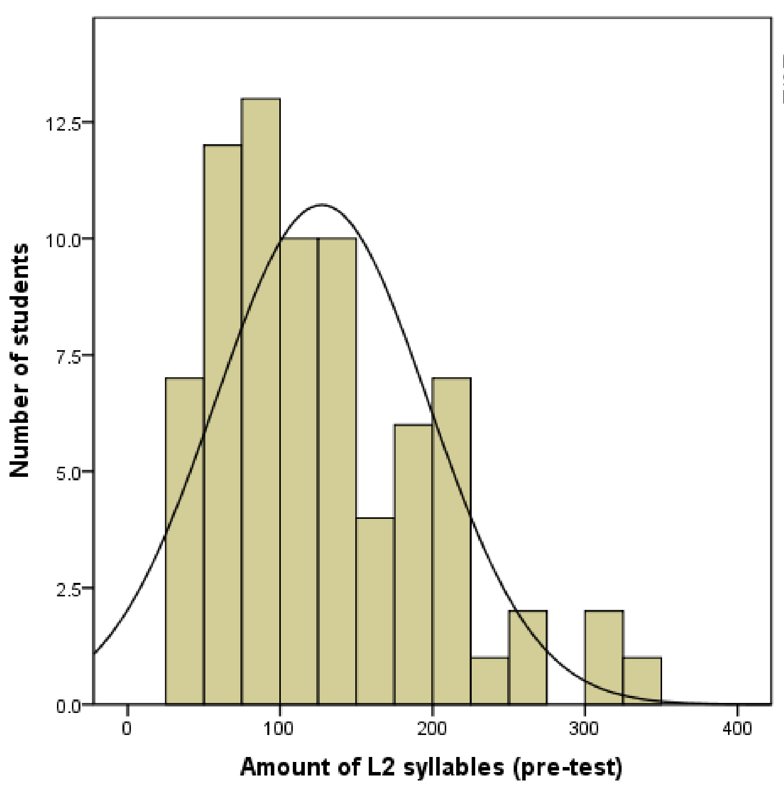


Skewness: 1.29

Kurtosis: 1.94

L2 syllables

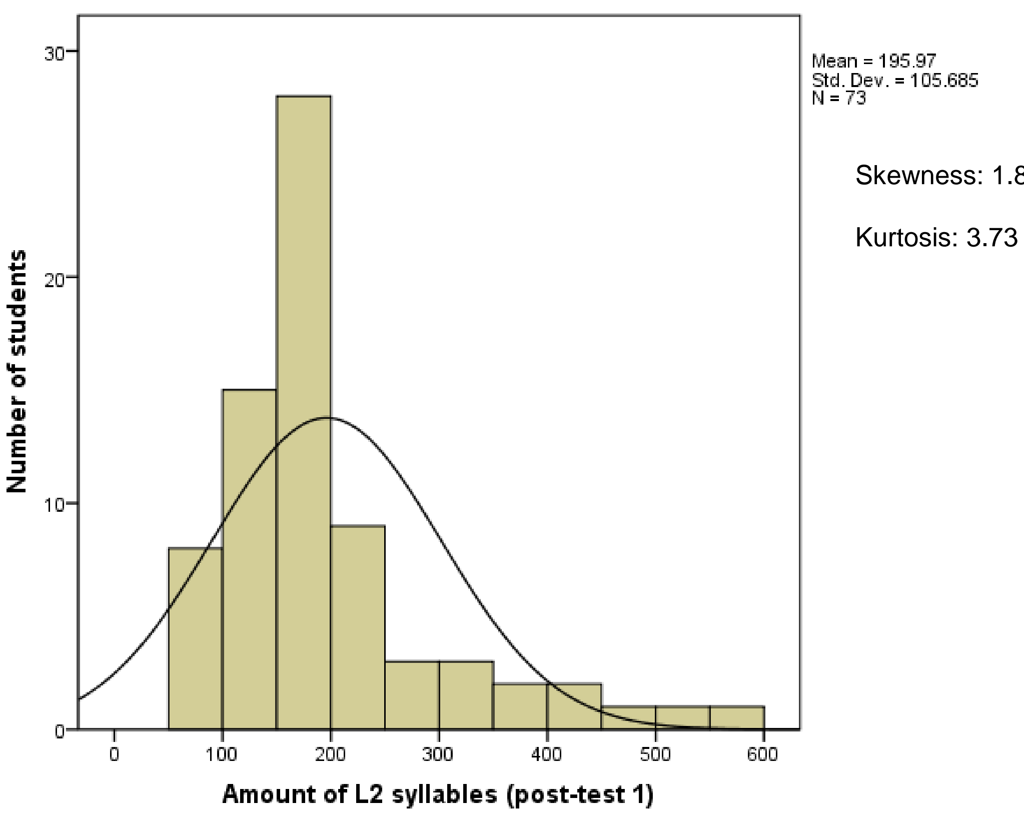
Pre-test



Skewness: .964

Kurtosis: .591

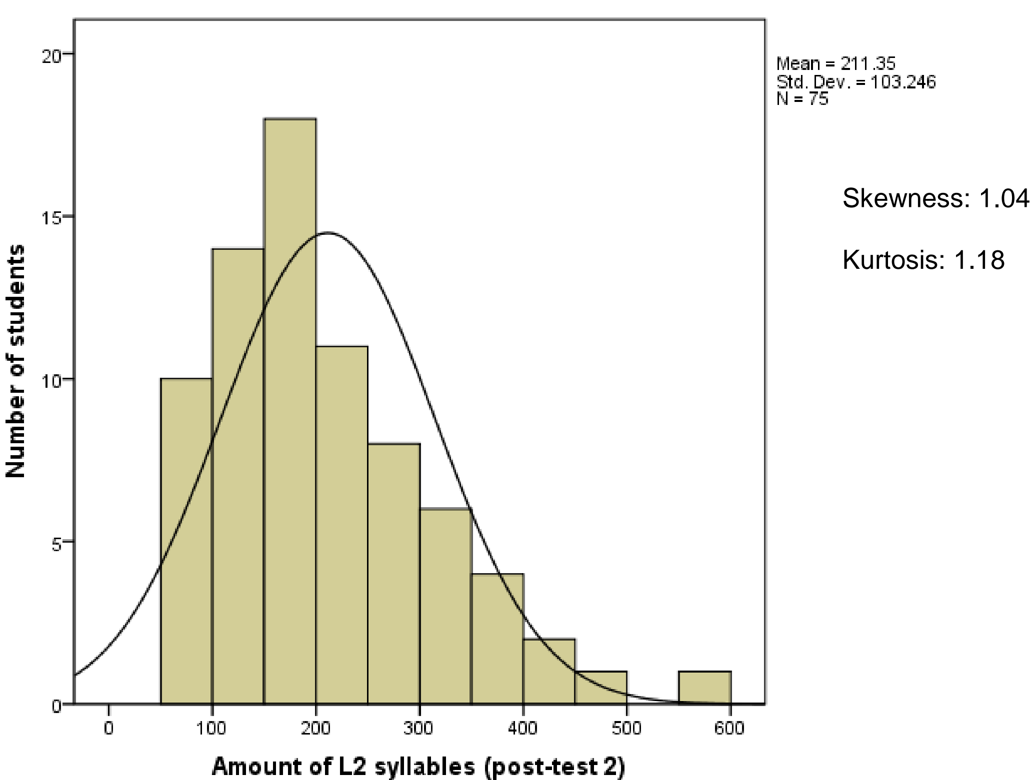
Post-test 1



Skewness: 1.81

Kurtosis: 3.73

Post-test 2



Skewness: 1.04

Kurtosis: 1.18

Appendix 37: Comparisons of results of parametric to non-parametric tests for WTC measures

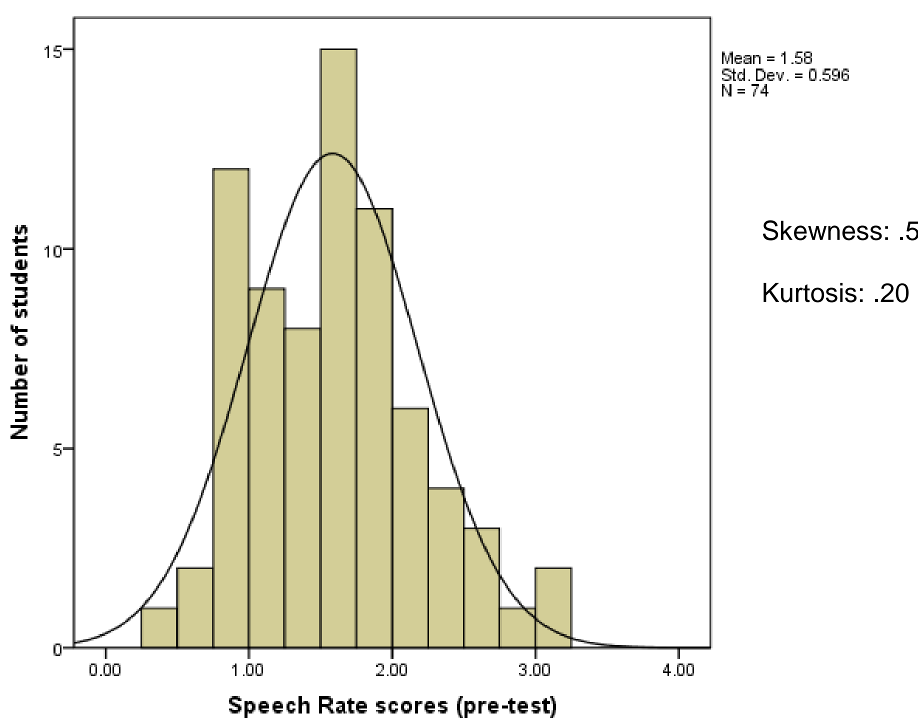
Speaking time	
Kruskal-Wallis	**ANOVA**
Pre-test: $U\,(2) = 2.88$, $p > .05$	**Pre-test:** $F\,(2, 73) = 1.39$, $p > .05$
Post-test 1: $U\,(2) = .11$, $p > .05$	**Post-test 1:** $F\,(2, 72) = .59$, $p > .05$
Post-test 2: $U\,(2) = 1.24$, $p > .05$	**Post-test 2:** $F\,(2, 74) = .51$, $p > .05$

L2 syllables	
Kruskal-Wallis	**ANOVA**
Pre-test: $U\,(2) = 6.68$, $p < .05$	**Pre-test:** $F\,(2, 74) = 3.11$, $p = .51$
Post-test 1: $U\,(2) = 10.69$, $p < .01$	**Post-test 1:** $F\,(2, 72) = 5.78$, $p < .01$
Post-test 2: $U\,(2) = 7.38$, $p < .05$	**Post-test 2:** $F\,(2, 74) = 3.87$, $p < .05$

Appendix 38: Distribution graphs for temporal measures of fluency

Speech Rate

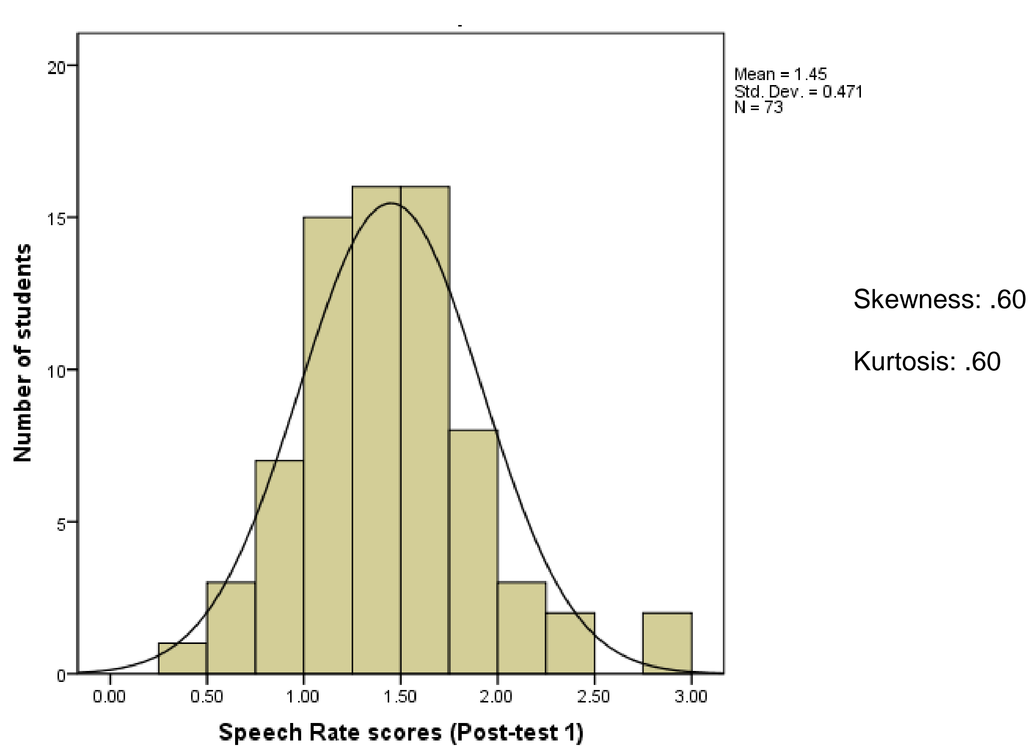
Pre-test



Skewness: .51

Kurtosis: .20

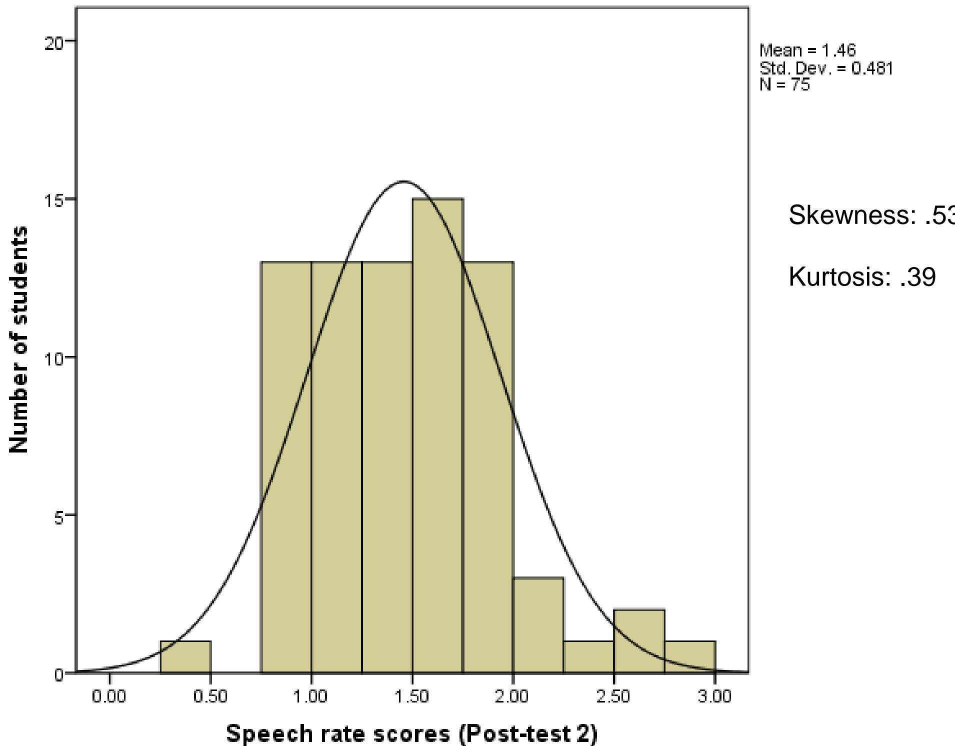
Post-test 1



Skewness: .60

Kurtosis: .60

Post-test 2

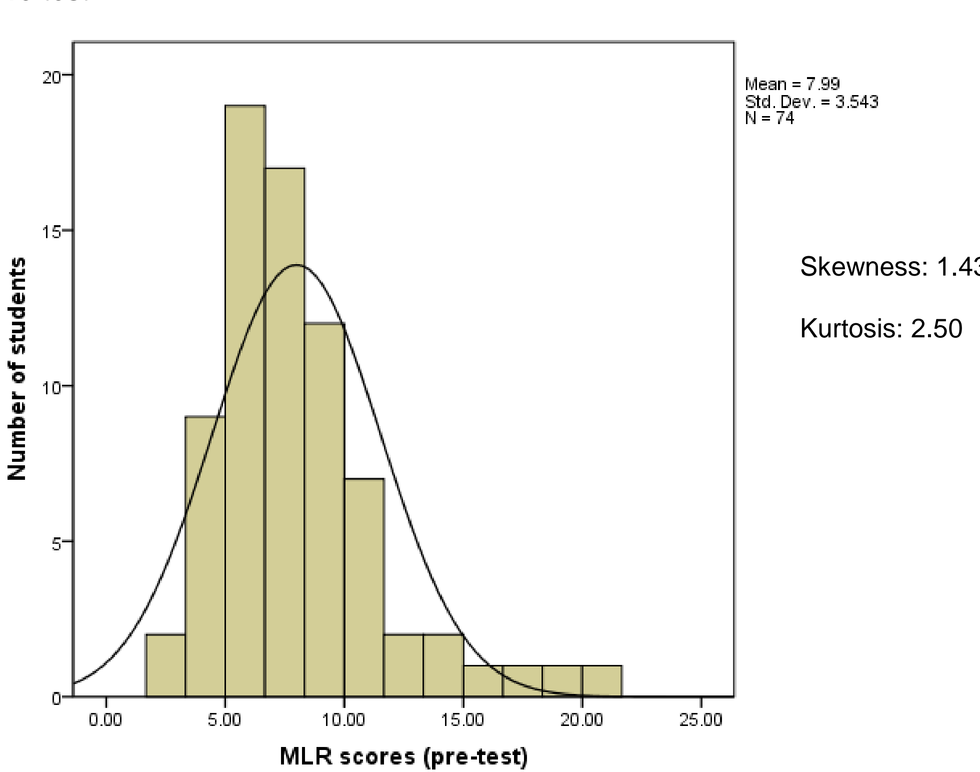


Skewness: .53

Kurtosis: .39

Mean Length of Runs

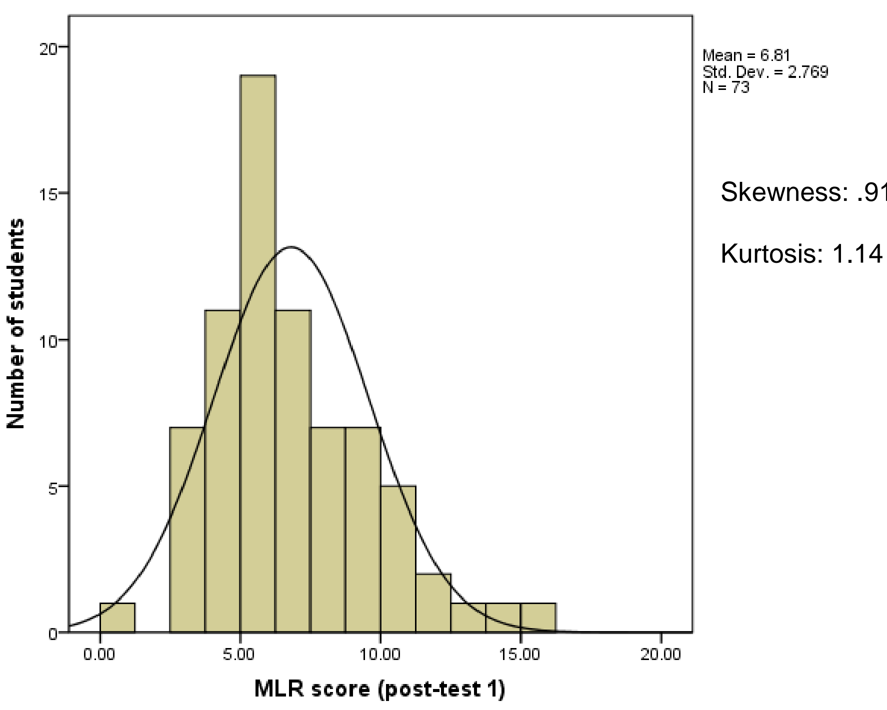
Pre-test



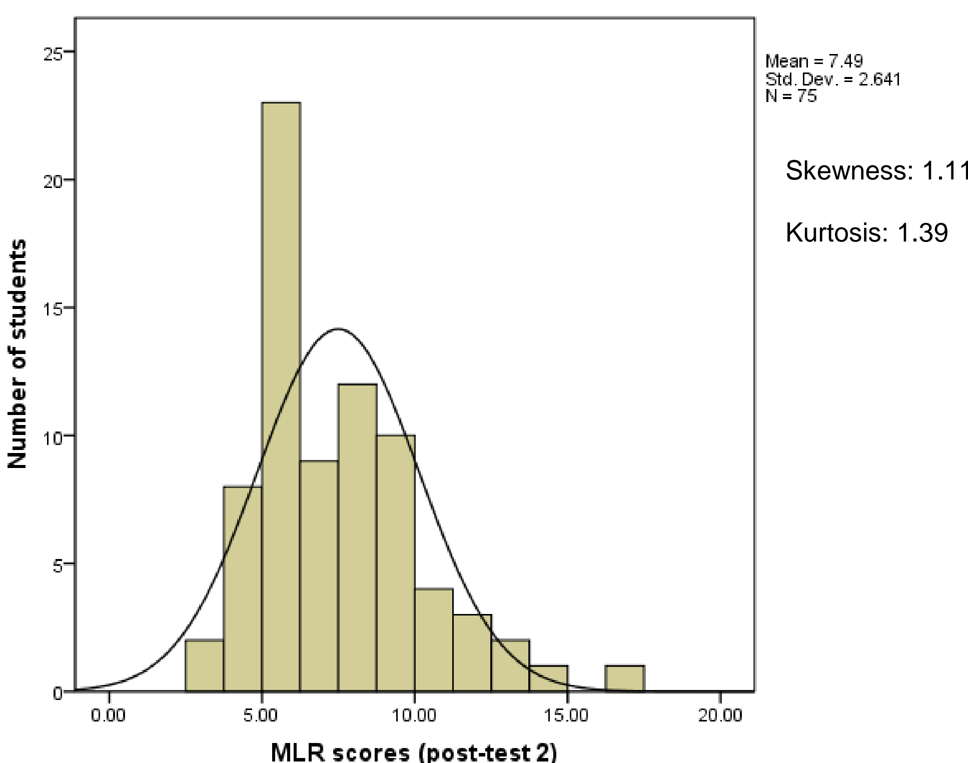
Skewness: 1.43

Kurtosis: 2.50

Post-test 1

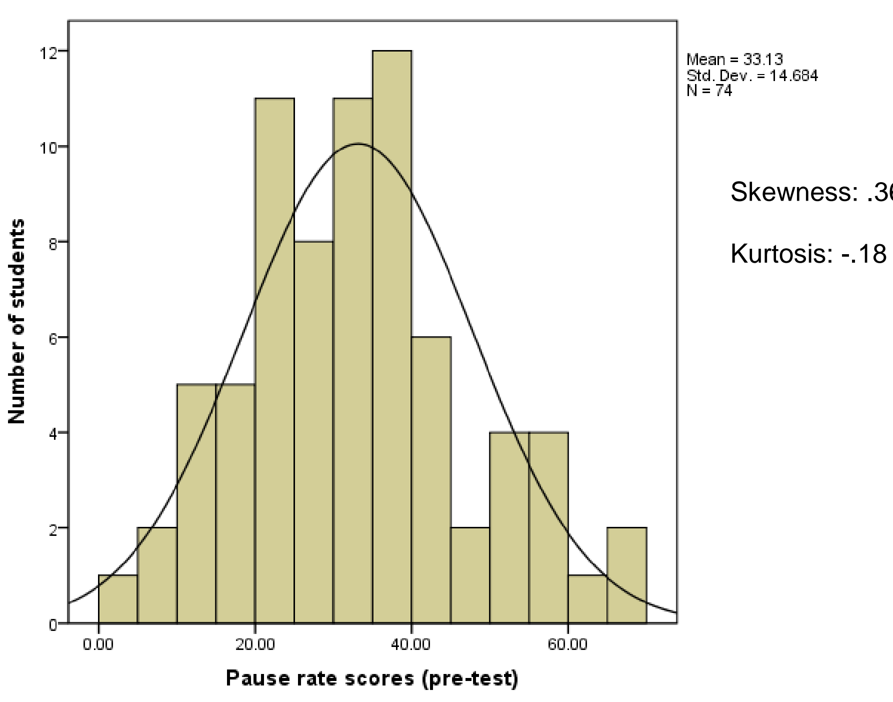


Post-test 2

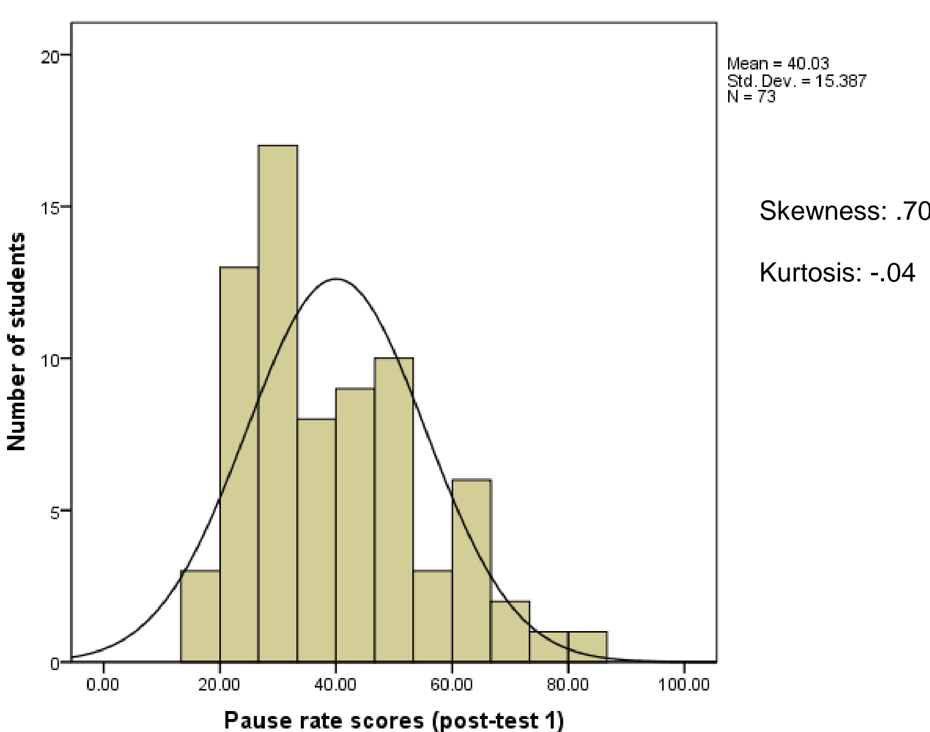


Pause Rate

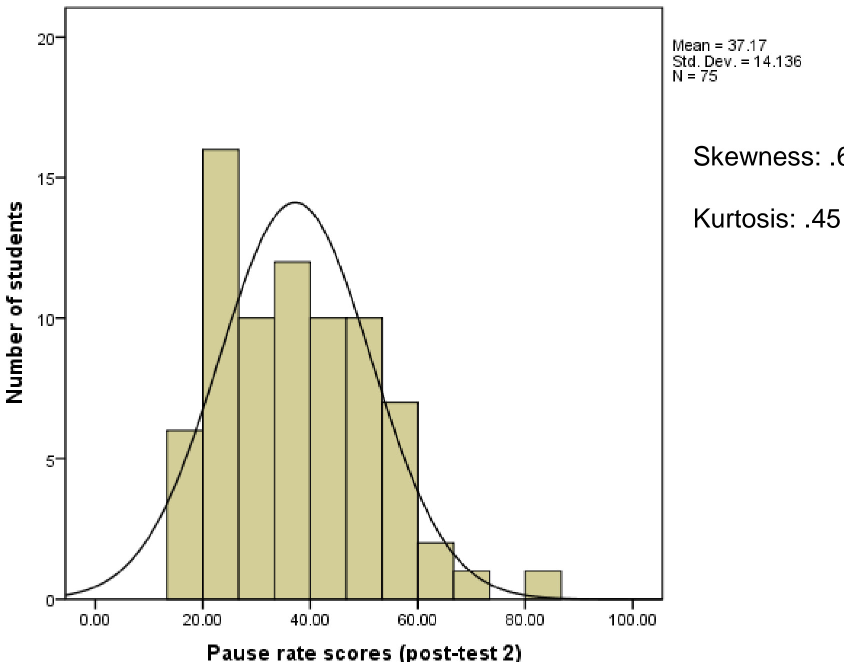
Pre-test



Post-test 1



Post-test 2



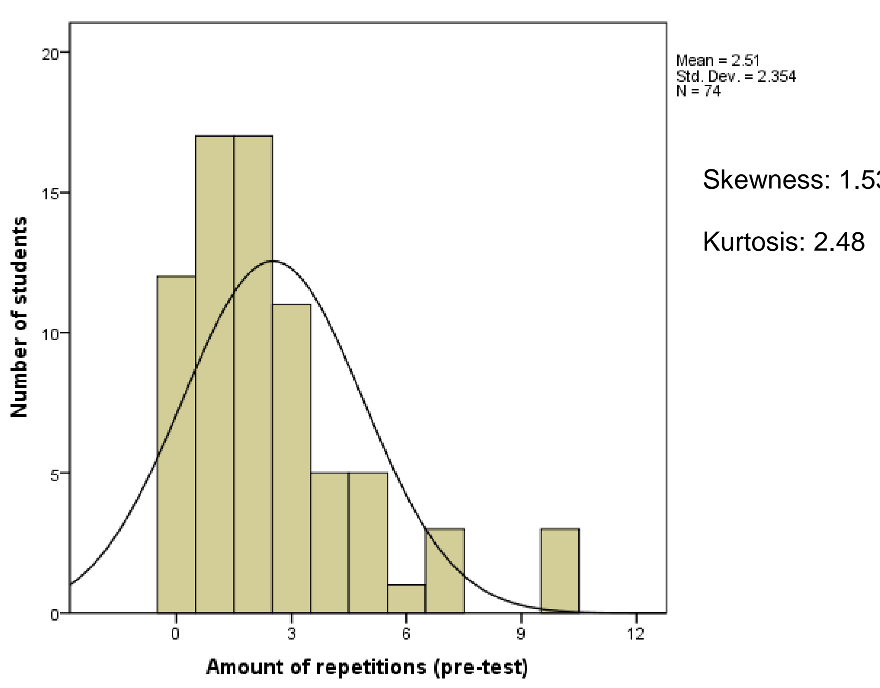
Skewness: .64

Kurtosis: .45

Appendix 39: Distribution graphs for dysfluency measures of fluency

Repetitions

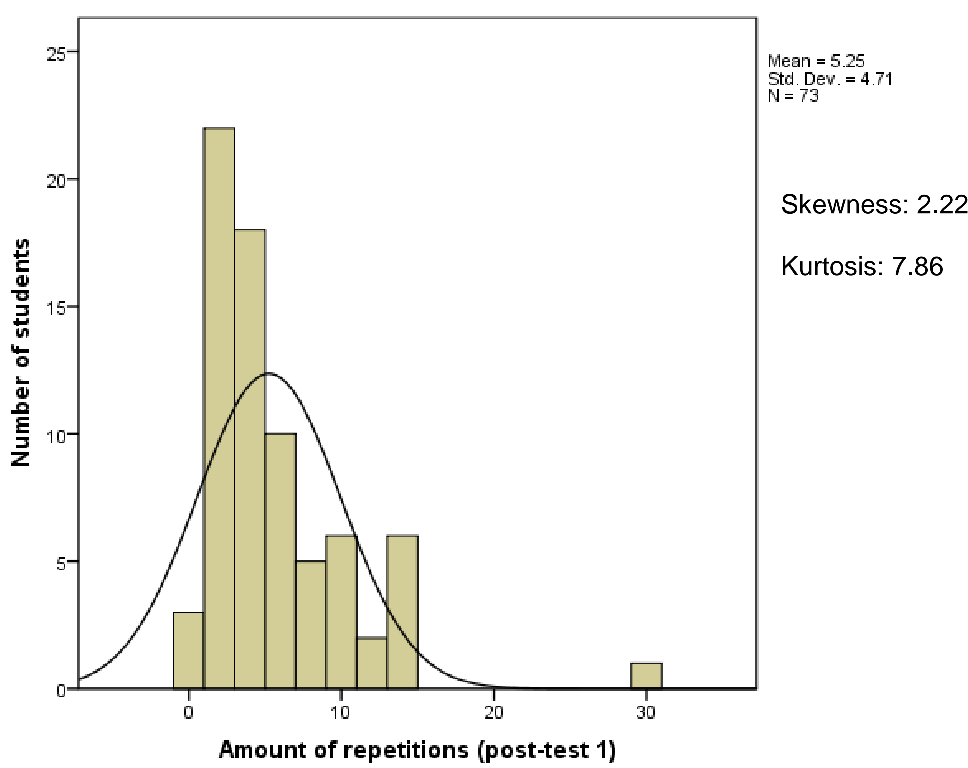
Pre-test



Skewness: 1.53

Kurtosis: 2.48

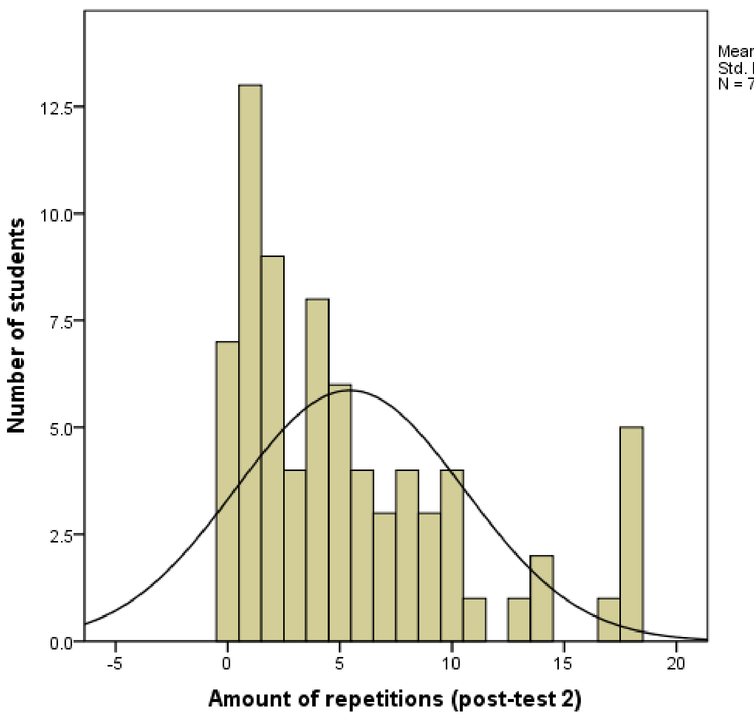
Post-test 1



Skewness: 2.22

Kurtosis: 7.86

Post-test 2

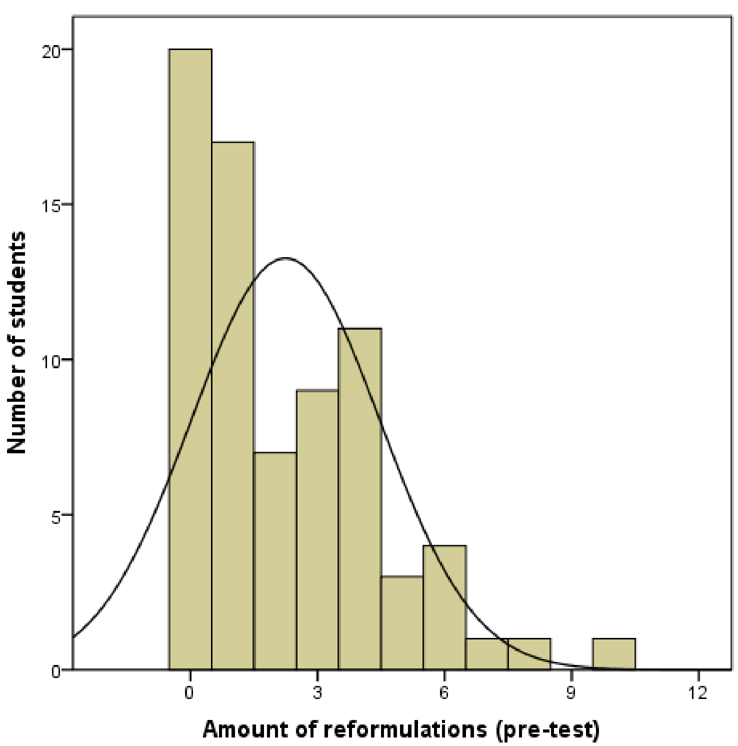


Skewness: 1.17

Kurtosis: .61

Reformulations

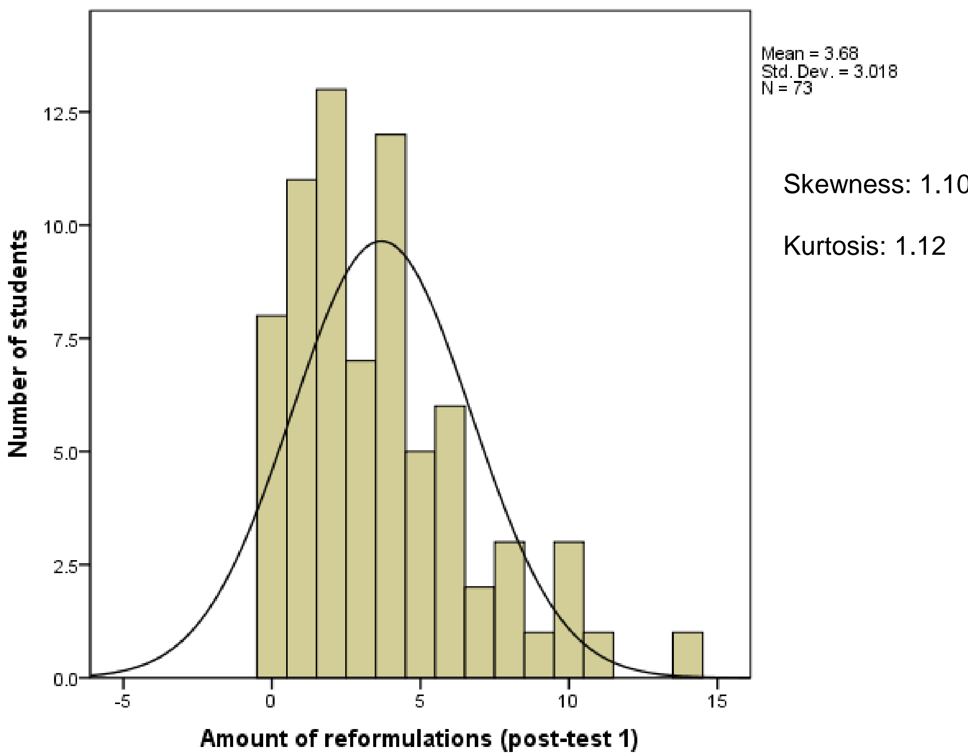
Pre-test



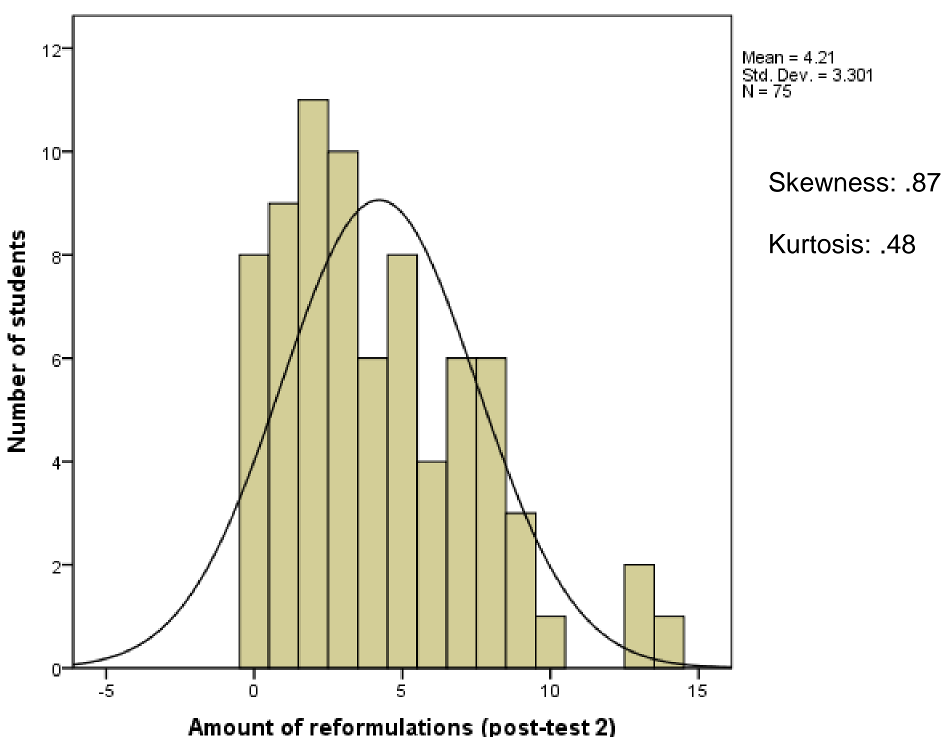
Skewness: 1.08

Kurtosis: 1.03

Post-test 1

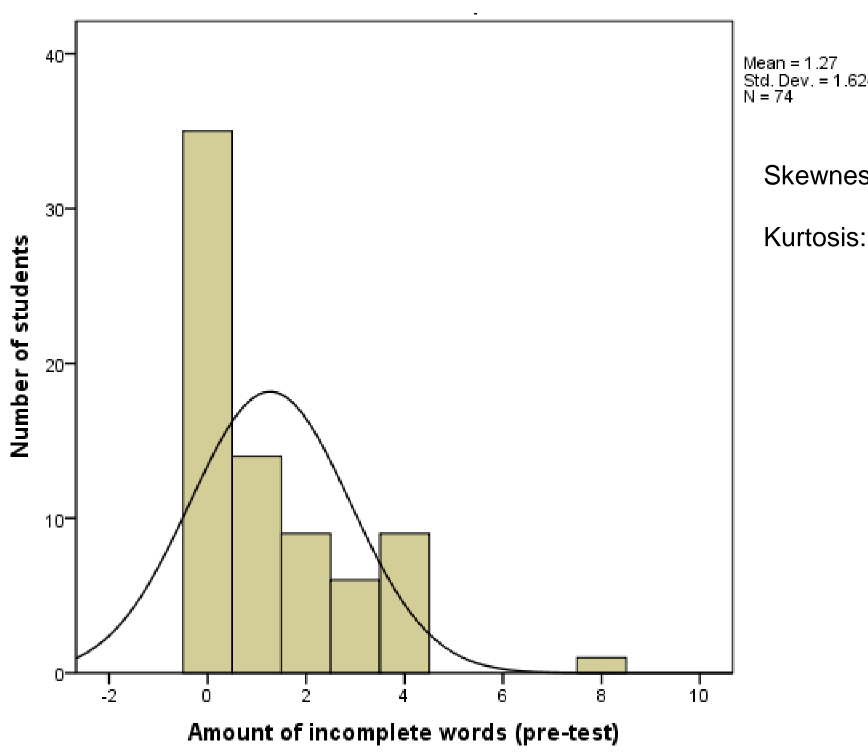


Post-test 2



Incomplete words

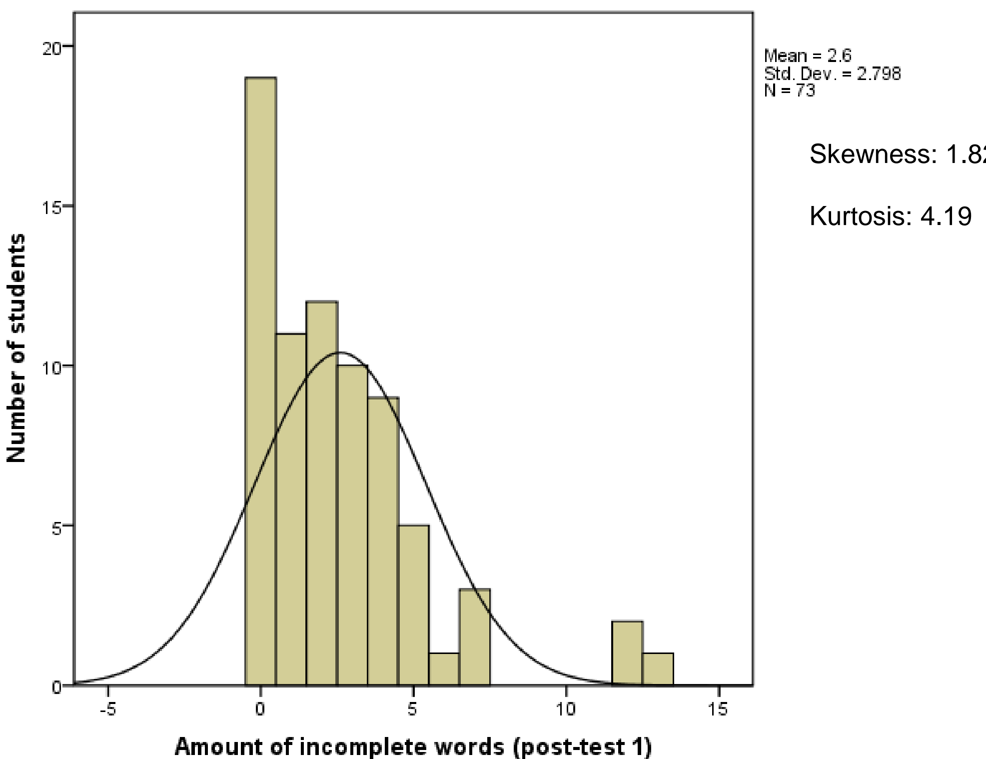
Pre-test



Skewness: 1.48

Kurtosis: 2.64

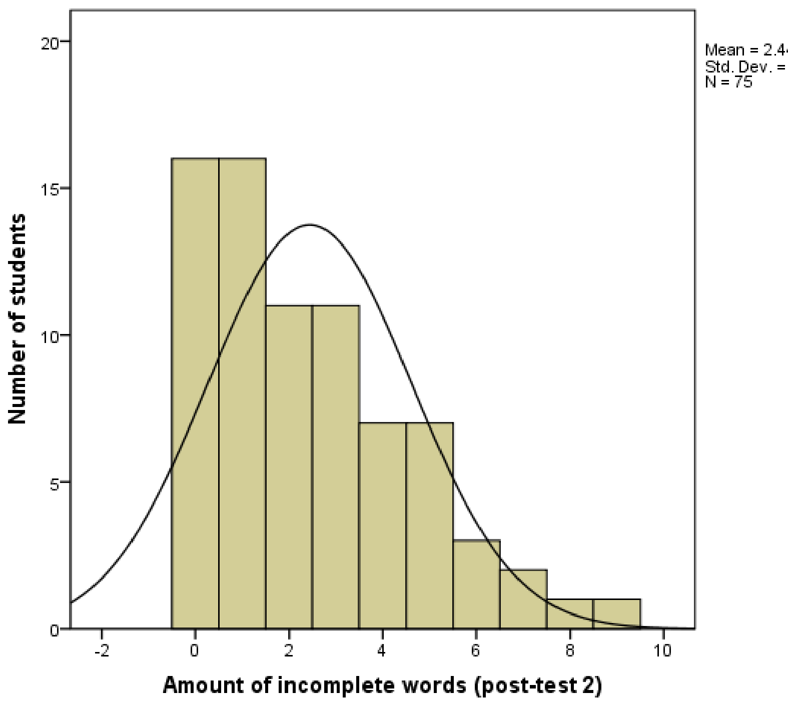
Post-test 1



Skewness: 1.82

Kurtosis: 4.19

Post-test 2



Skewness: .87

Kurtosis: .23

Appendix 40: Comparisons of results of parametric to non-parametric tests for all fluency measures.

Mean Length of Runs	
Kruskal-Wallis	**ANOVA**
Pre-test: U (2) = 5.15, p = .076	**Pre-test:** F (2, 73) = 3.08, p = .052
Post-test 1: U (2) = 7.81, $p < .05$	**Post-test 1:** F (2, 72) = 4.19, $p < .05$
Post-test 2: U (2) = 10.89, $p < .01$	**Post-test 2:** F (2, 74) = 5.15, $p < .01$

Pause rate	
Kruskal-Wallis	**ANOVA**
Post-test 1: U (2) = 9.85, $p < .01$	**Post-test 1:** F (2, 72) = 6.46, $p < .05$

Repetitions	
Kruskal-Wallis	**ANOVA**
Pre-test: U (2) = 2.79, $p > .05$	**Pre-test:** F (2, 73) = 1.91, $p > .05$
Post-test 1: U (2) = .19, $p > .05$	**Post-test 1:** F (2, 72) = .64, $p > .05$
Post-test 2: U (2) = .67, $p > .05$	**Post-test 2:** F (2, 74) = .27, $p > .05$

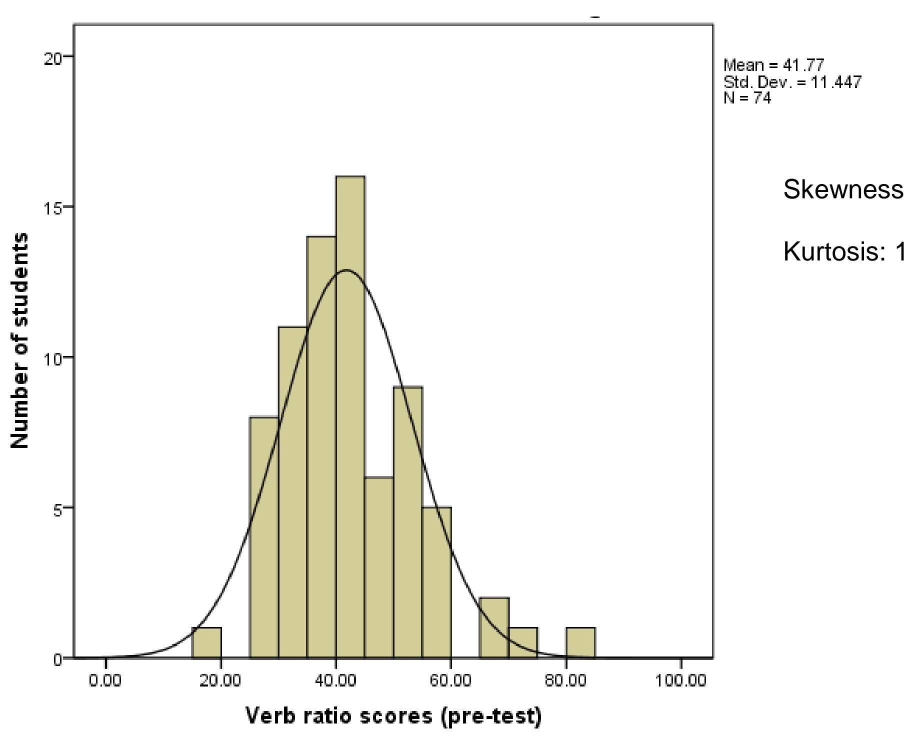
Reformulations	
Kruskal-Wallis	**ANOVA**
Pre-test: U (2) = 2.67, $p > .05$	**Pre-test:** F (2, 73) = 1.59, $p > .05$
Post-test 1: U (2) = 1.71, $p > .05$	**Post-test 1:** F (2, 72) = .54, $p > .05$
Post-test 2: U (2) = .44, $p > .05$	**Post-test 2:** F (2, 74) = .08, $p > .05$

Incomplete words	
Kruskal-Wallis	**ANOVA**
Pre-test: U (2) = 3.21, $p > .05$	**Pre-test:** F (2, 73) = .63, $p > .05$
Post-test 1: U (2) = .003, $p > .05$	**Post-test 1:** F (2, 72) = .81, $p > .05$
Post-test 2: U (2) = 1.11, $p > .05$	**Post-test 2:** F (2, 74) = .35, $p > .05$

Appendix 41: Distribution graphs for content measures

Verb Ratio

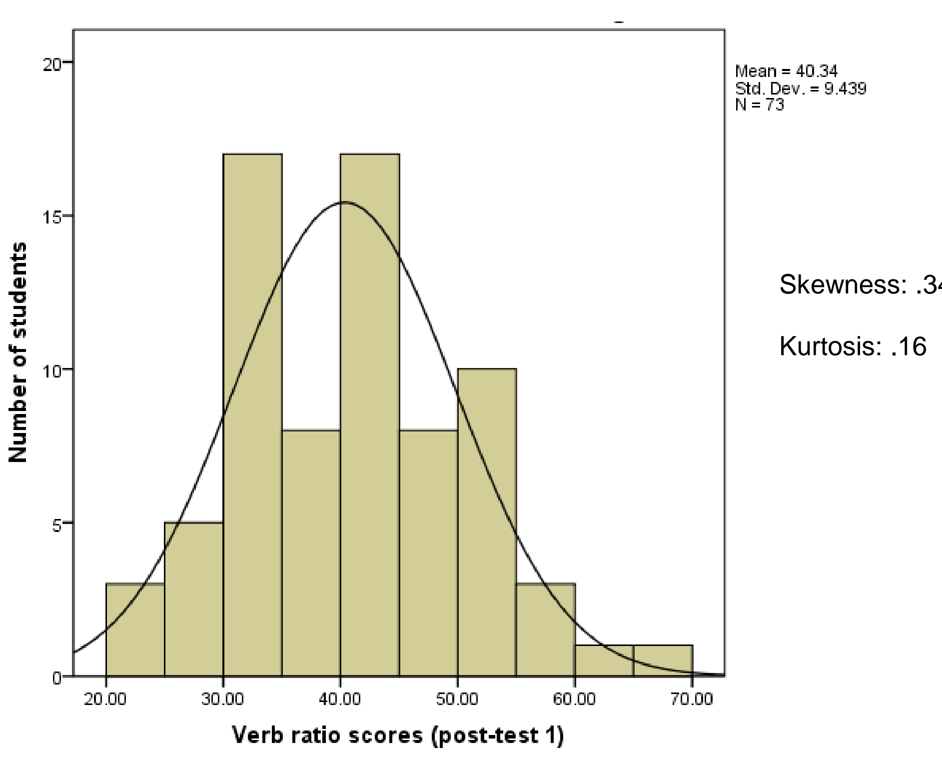
Pre-test



Skewness: .94

Kurtosis: 1.74

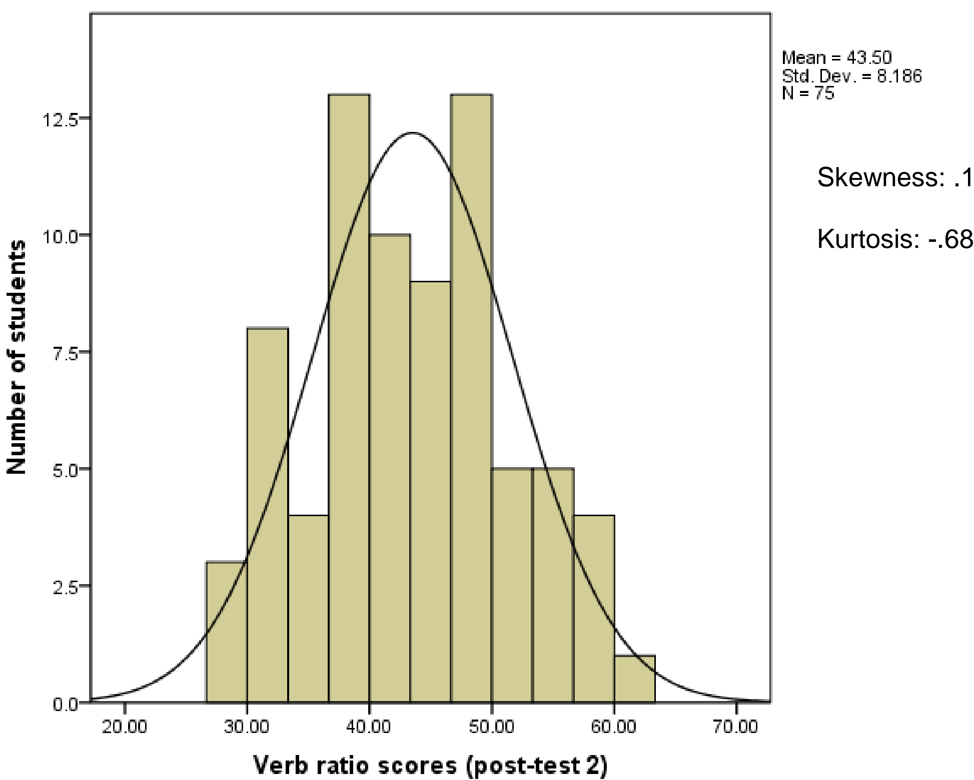
Post-test 1



Skewness: .34

Kurtosis: .16

Post-test 2

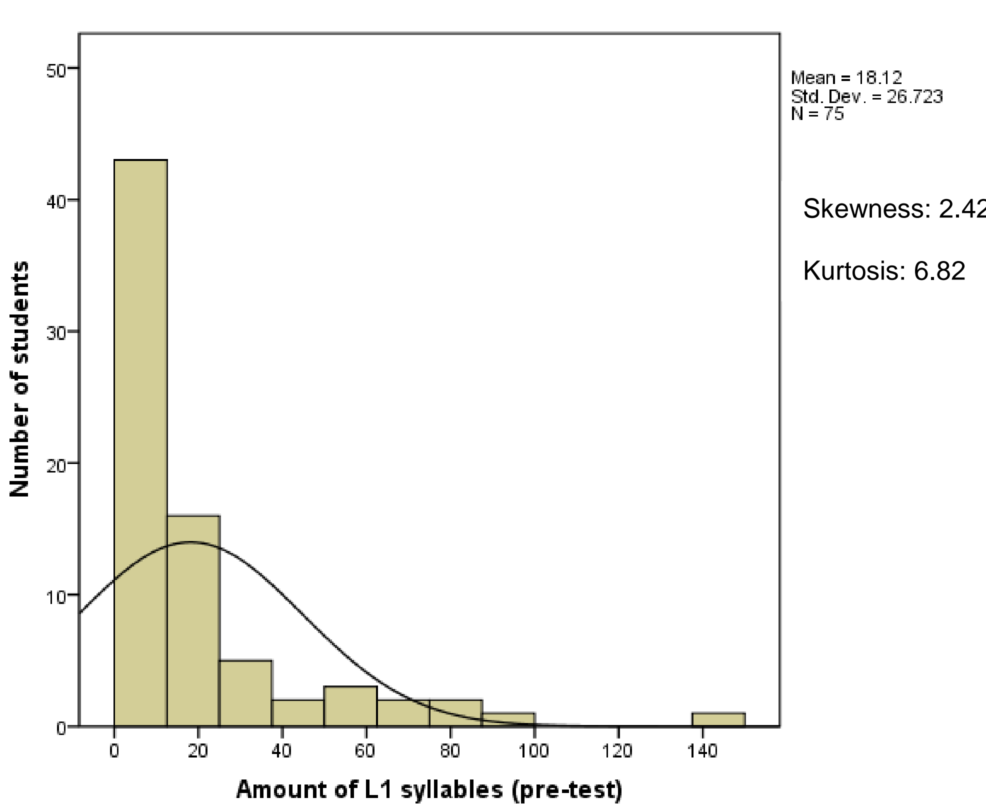


Skewness: .15

Kurtosis: -.68

L1 syllables

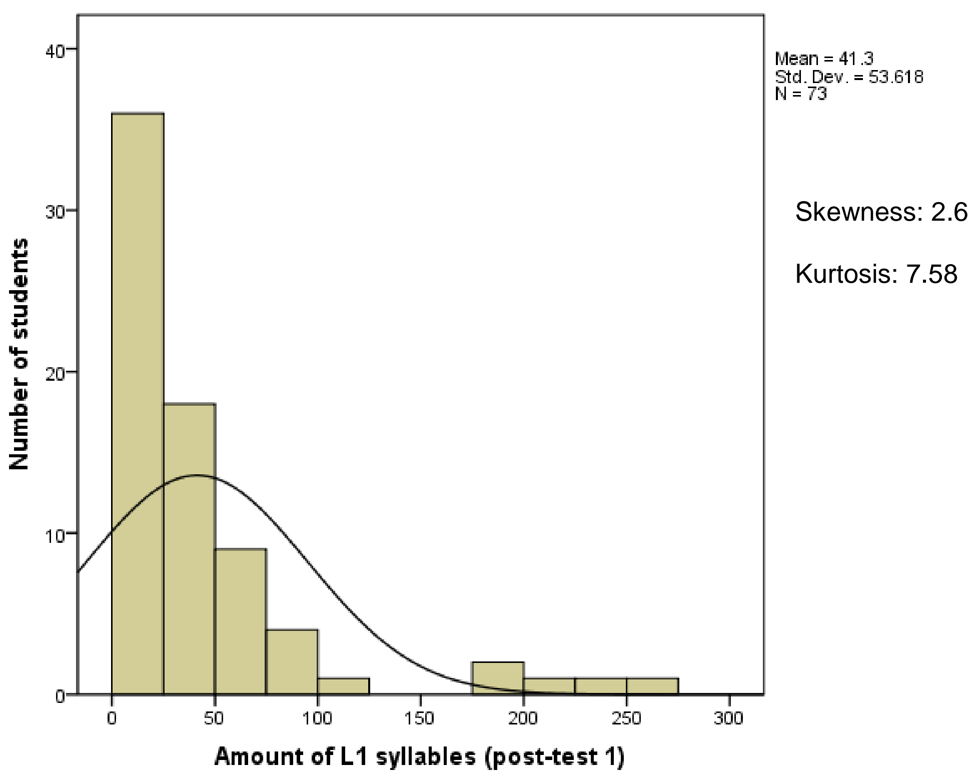
Pre-test



Skewness: 2.42

Kurtosis: 6.82

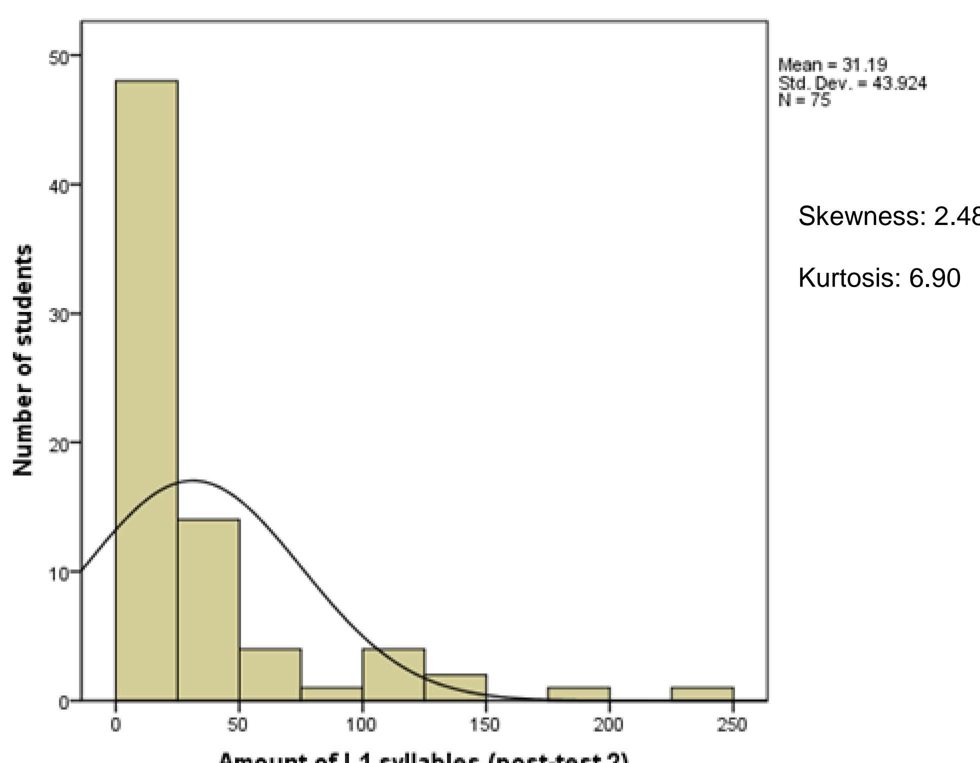
Post-test 1



Skewness: 2.67

Kurtosis: 7.58

Post-test 2

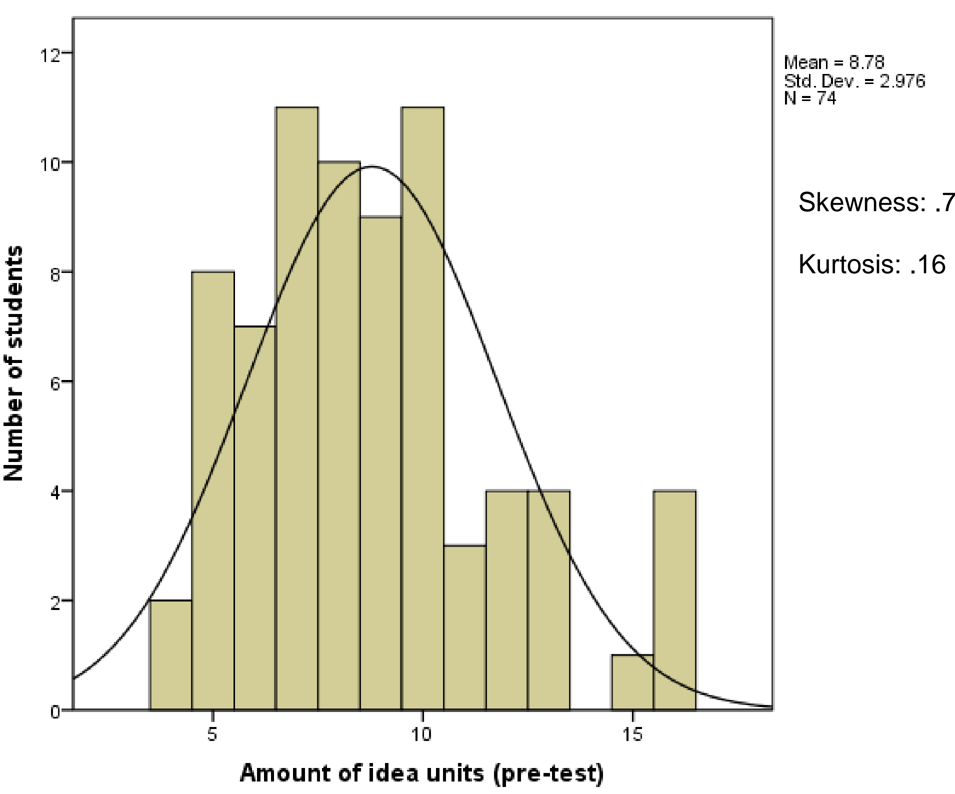


Skewness: 2.48

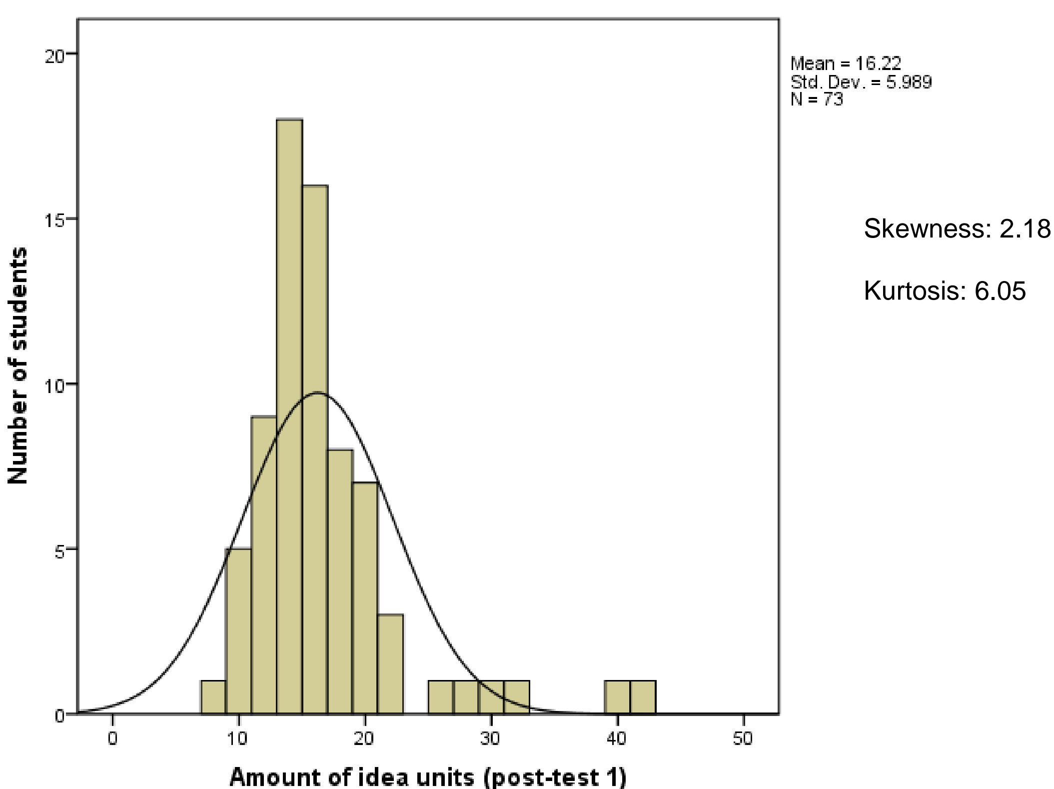
Kurtosis: 6.90

Idea units

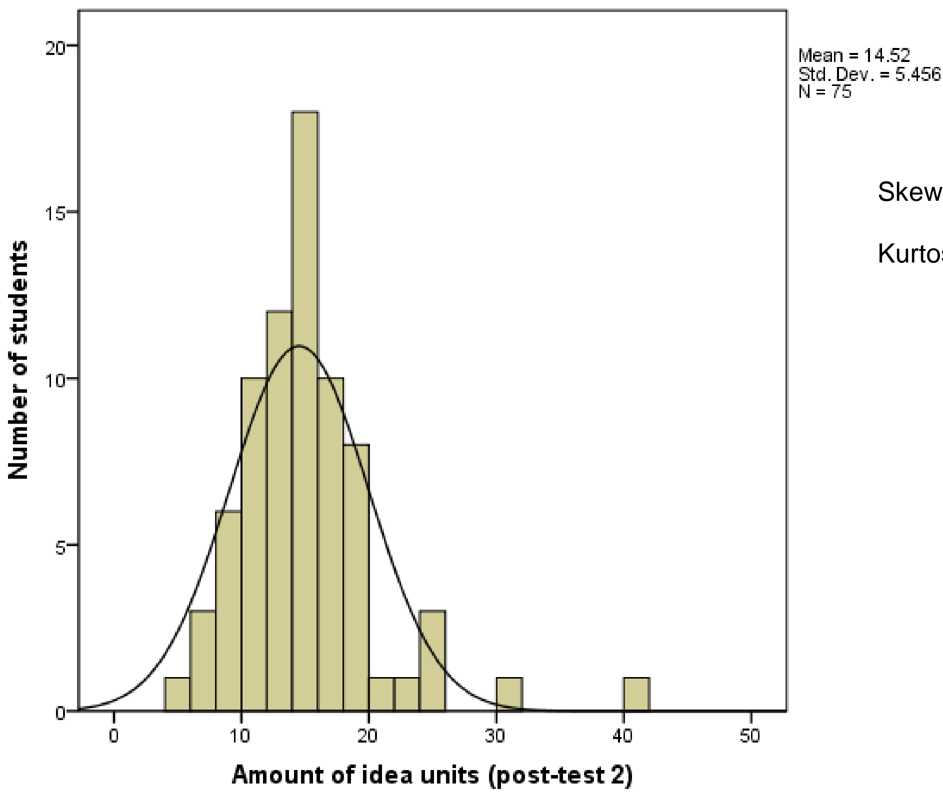
Pre-test



Post-test 1



Post-test 2



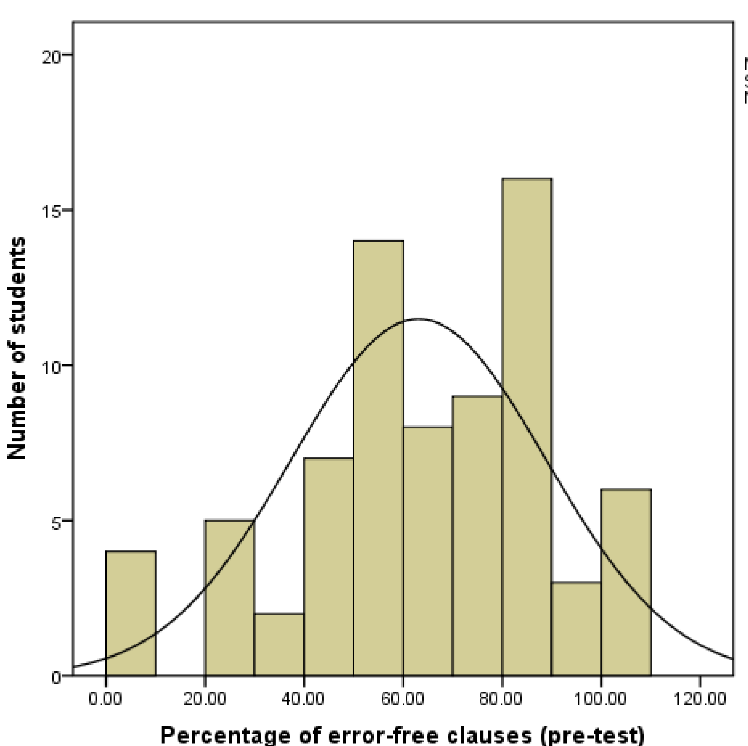
Skewness: 1.80

Kurtosis: 6.30

Appendix 42: Distribution graphs for accuracy measures

Percentage of error-free clauses

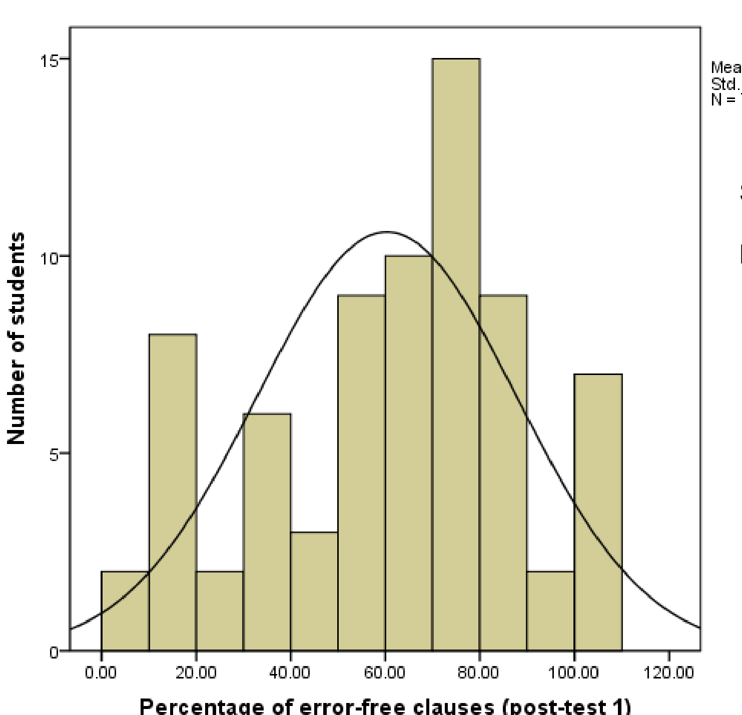
Pre-test



Skewness: -.65

Kurtosis: -.08

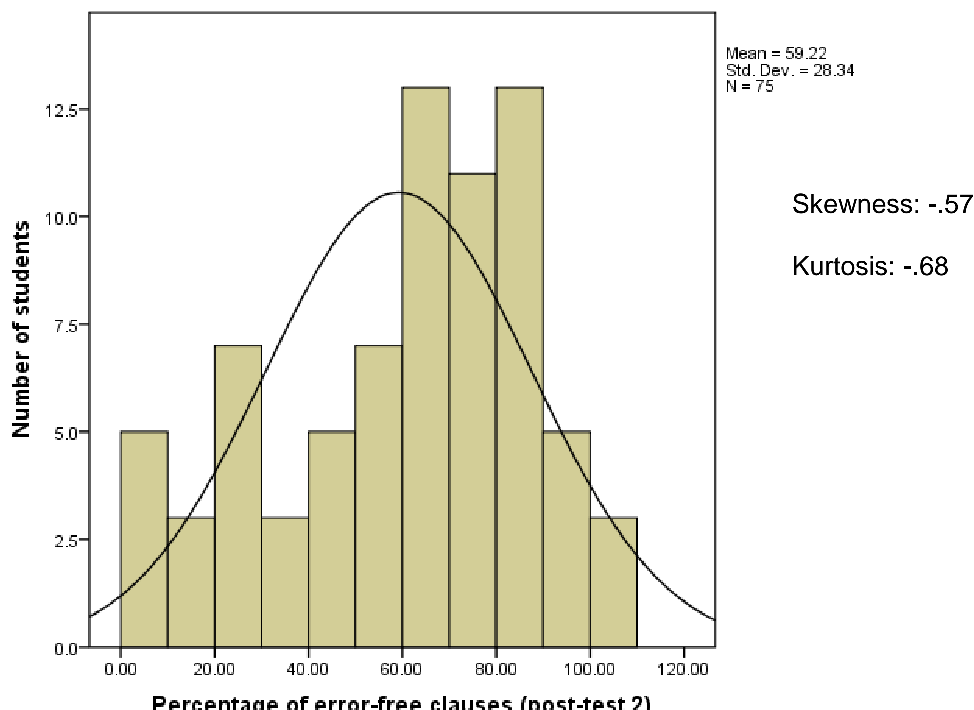
Post-test 1



Skewness: -.50

Kurtosis: -.67

Post-test 2

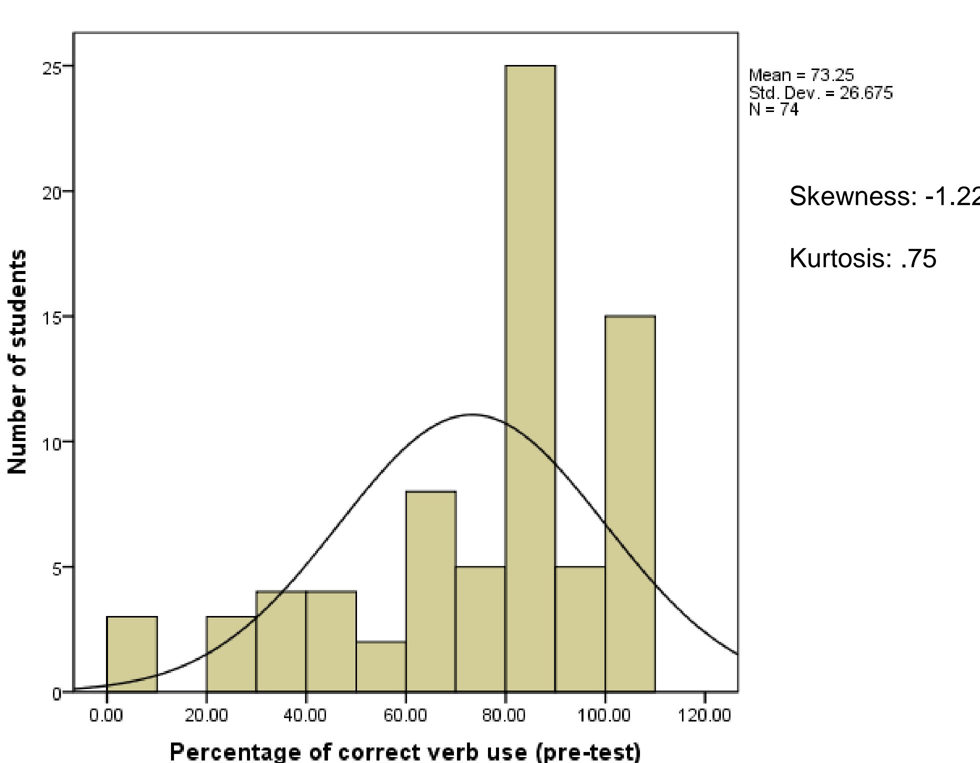


Skewness: -.57

Kurtosis: -.68

Percentage of correct verb use

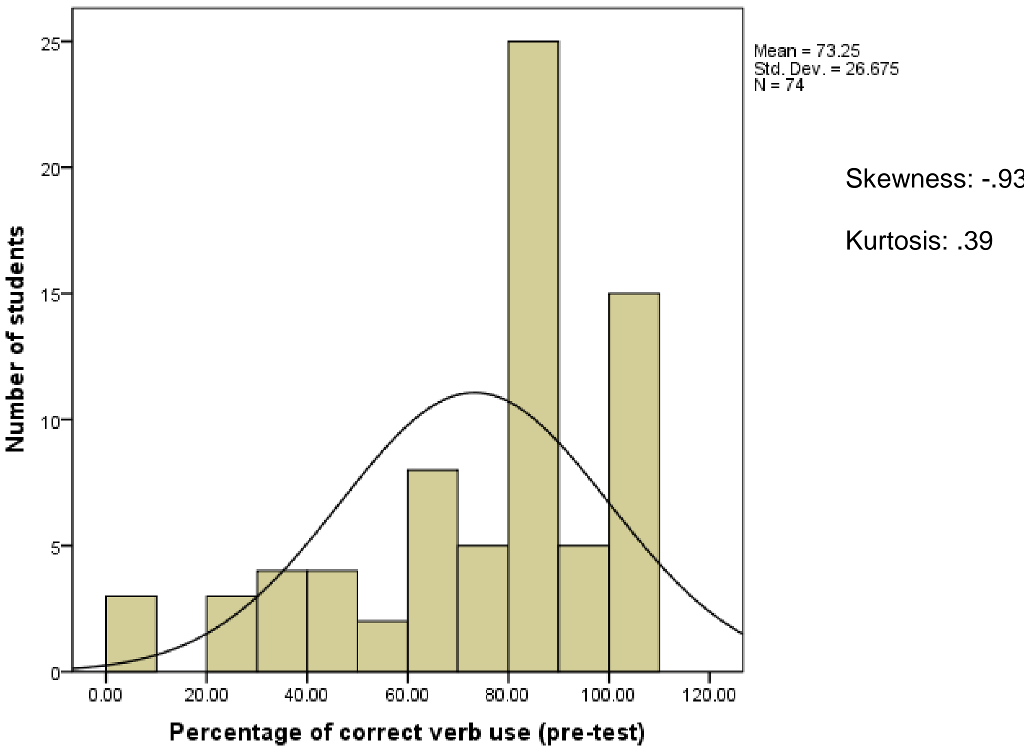
Pre-test



Skewness: -1.22

Kurtosis: .75

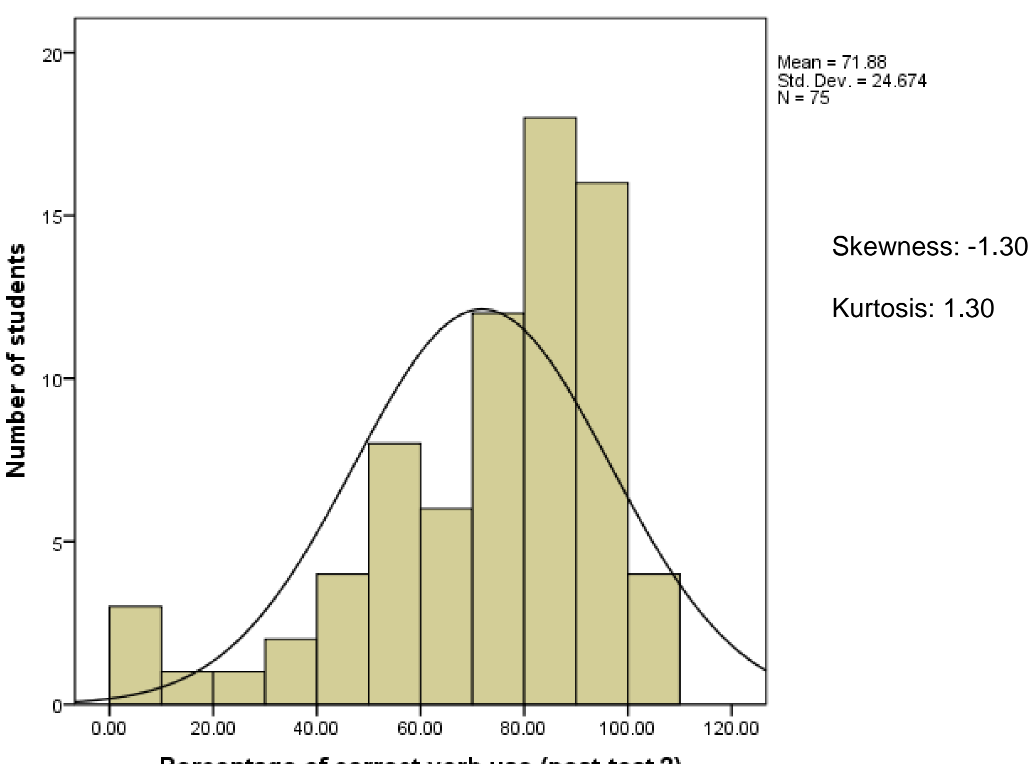
Post-test 1



Skewness: -.93

Kurtosis: .39

Post-test 2



Skewness: -1.30

Kurtosis: 1.30

Appendix 43: Comparisons of results of parametric to non-parametric tests for content and accuracy measures.

L1 syllables	
Kruskal-Wallis	**ANOVA**
Pre-test: U (2) = 4.95, $p > .05$	**Pre-test:** F (2, 74) = 2.29, $p > .05$
Post-test 1: U (2) = 5.60, $p > .05$	**Post-test 1:** F (2, 72) = 2.61, $p > .05$
Post-test 2: U (2) = 3.72, $p > .05$	**Post-test 2:** F (2, 74) = 2.67, $p > .05$

Idea Units	
Kruskal-Wallis	**ANOVA**
Pre-test: U (2) = 3.59, $p > .05$	**Pre-test:** F (2, 73) = 2.27, $p > .05$
Post-test 1: U (2) = 4.08, $p > .05$	**Post-test 1:** F (2, 72) = 3.88, $p < .05$
Post-test 2: U (2) = 5.24, $p > .05$	**Post-test 2:** F (2, 74) = 2.08, $p > .05$

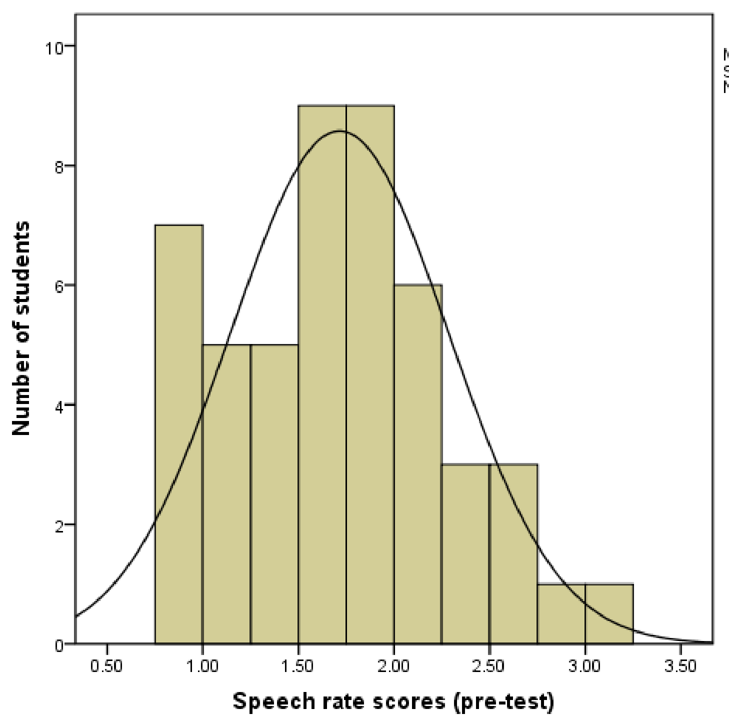
Percentage of error-free clauses	
Kruskal-Wallis	**ANOVA**
Post-test 1: U (2) = 3.72, $p > .05$	**Post-test 1:** F (2, 72) = 1.61, $p > .05$
Post-test 2: U (2) = 13.75, $p = .001$	**Post-test 2:** F (2, 74) = 9.17, $p < .001$

Percentage of correct verb use	
Kruskal-Wallis	**ANOVA**
Pre-test: U (2) = 6.43, $p < .05$	**Pre-test:** F (2, 73) = 6.39, $p < .05$
Post-test 1: U (2) = 7.49, $p < .05$	**Post-test 1:** F (2, 72) = 3.55, $p < .05$
Post-test 2: U (2) = 16.33, $p < .01$	**Post-test 2:** F (2, 74) = 7.56, $p = .001$

Appendix 44: Distribution graphs for temporal variables of fluency (for proficiency groups)

Speech Rate

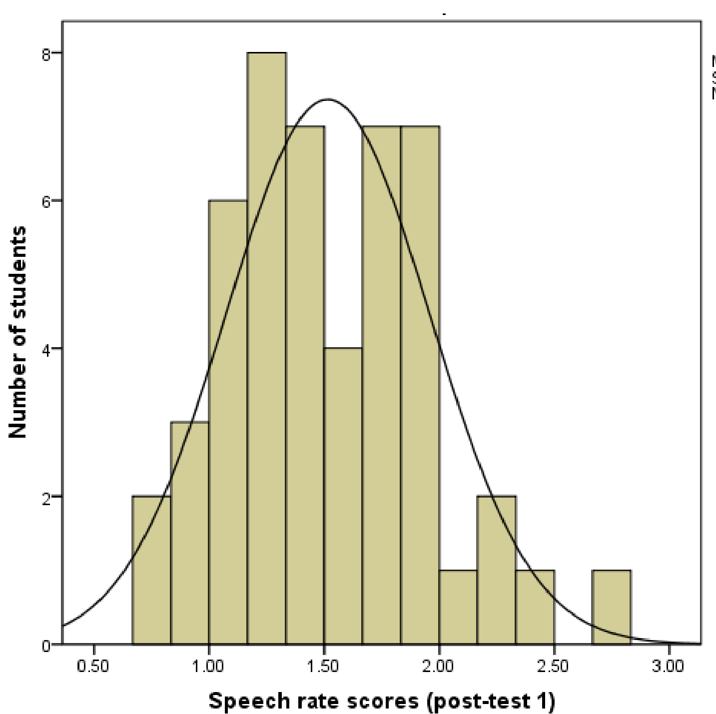
Pre-test



Skewness: .47

Kurtosis: -.02

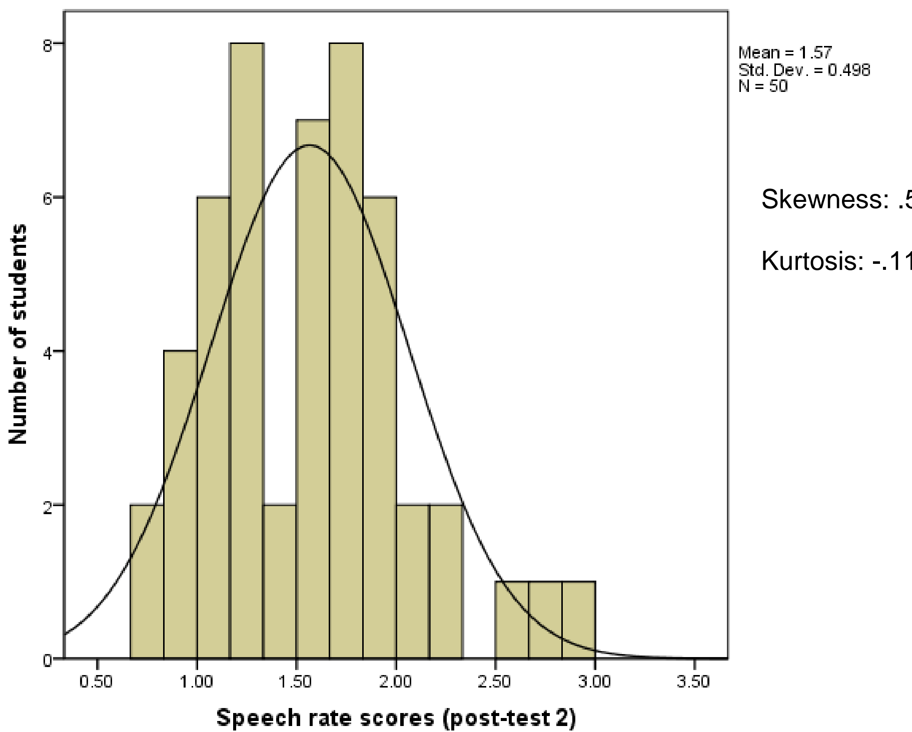
Post-test 1



Skewness: .43

Kurtosis: .19

Post-test 2

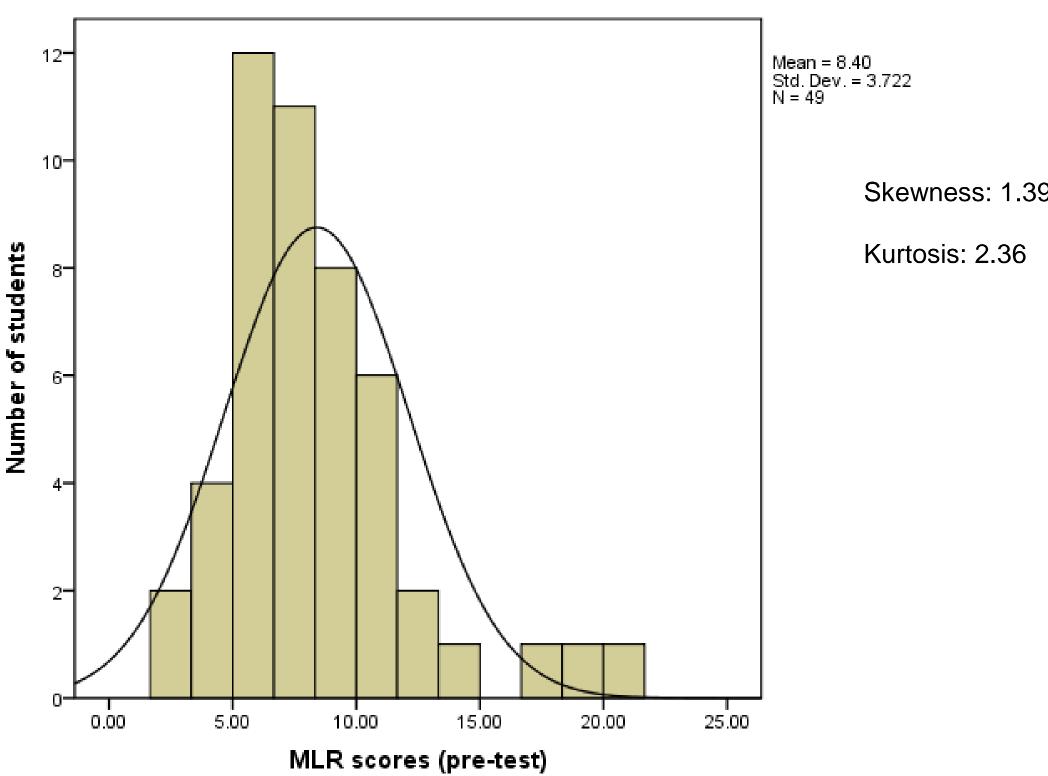


Skewness: .52

Kurtosis: -.11

Mean Length of Runs

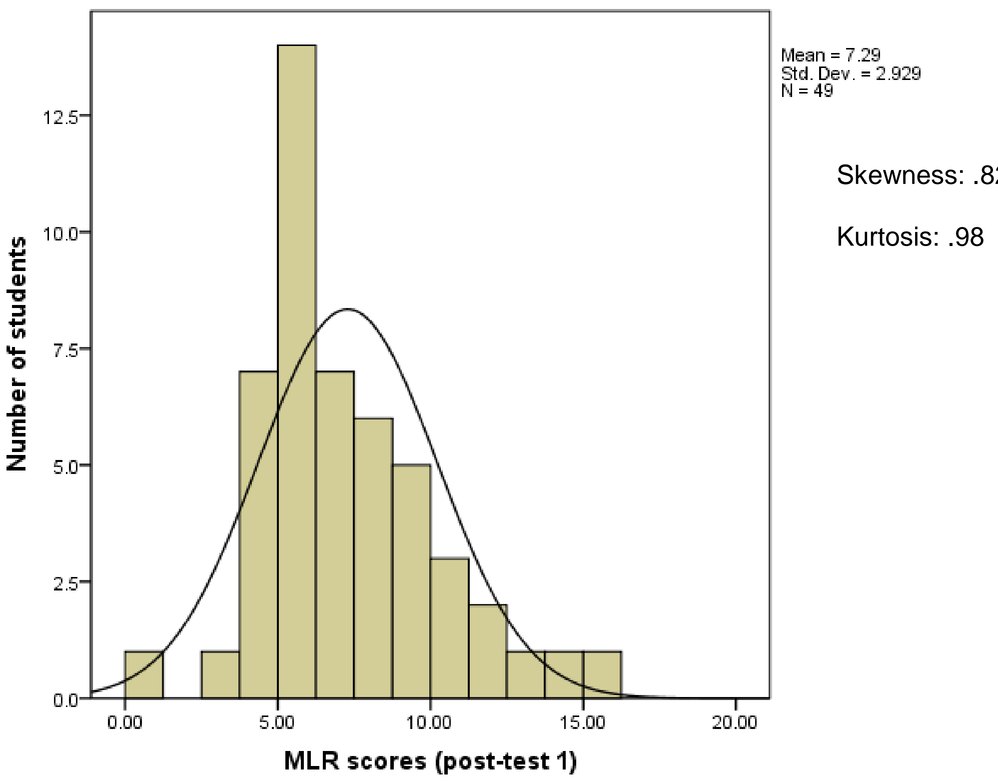
Pre-test



Skewness: 1.39

Kurtosis: 2.36

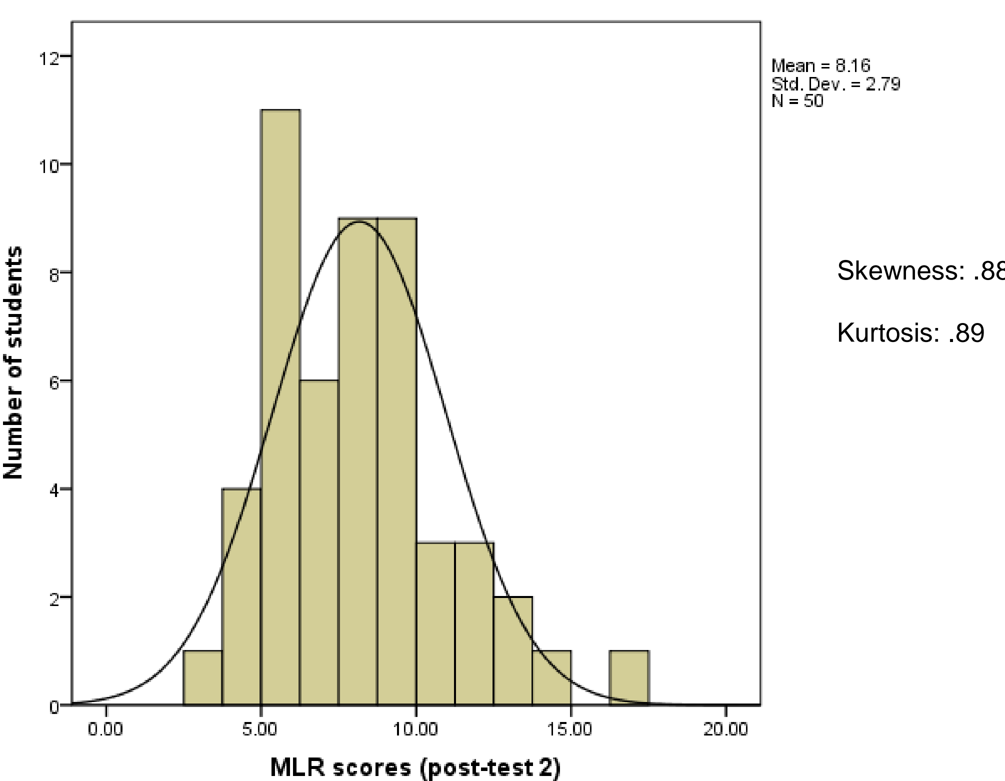
Post-test 1



Skewness: .82

Kurtosis: .98

Post-test 2

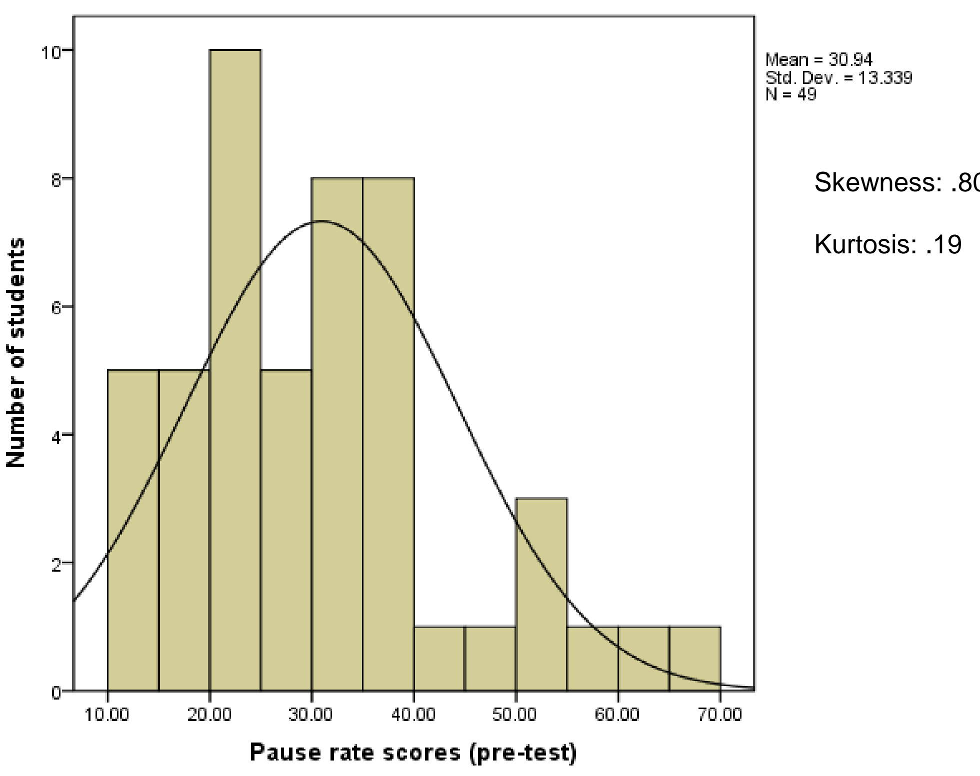


Skewness: .88

Kurtosis: .89

Pause Rate

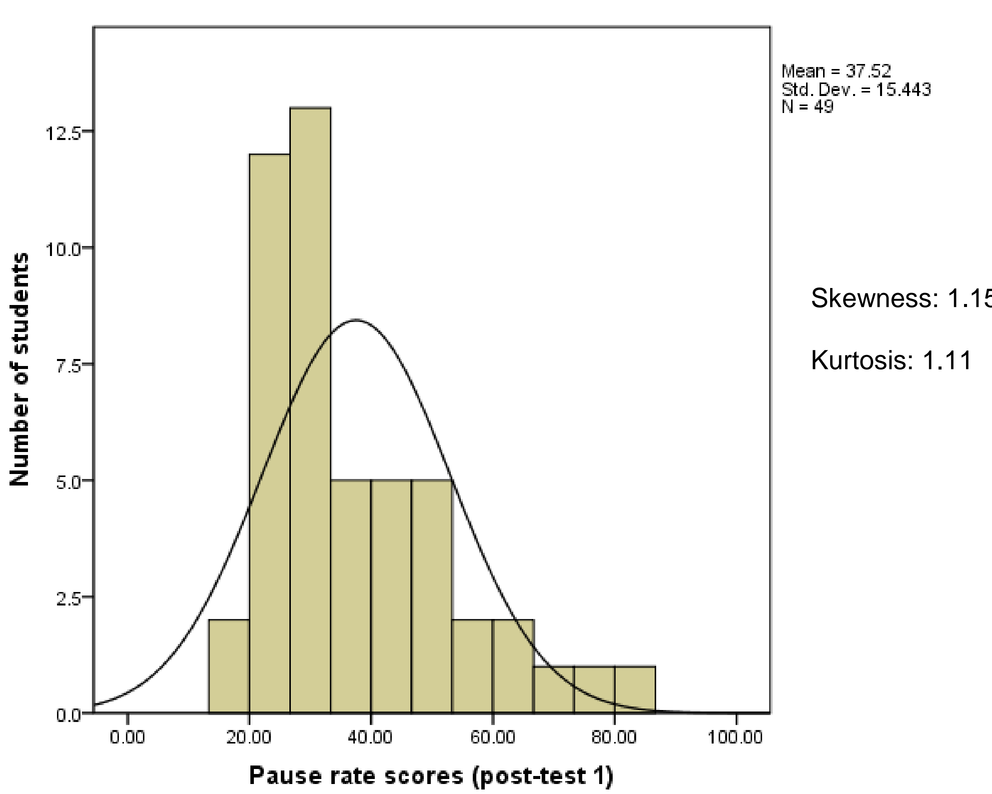
Pre-test



Skewness: .80

Kurtosis: .19

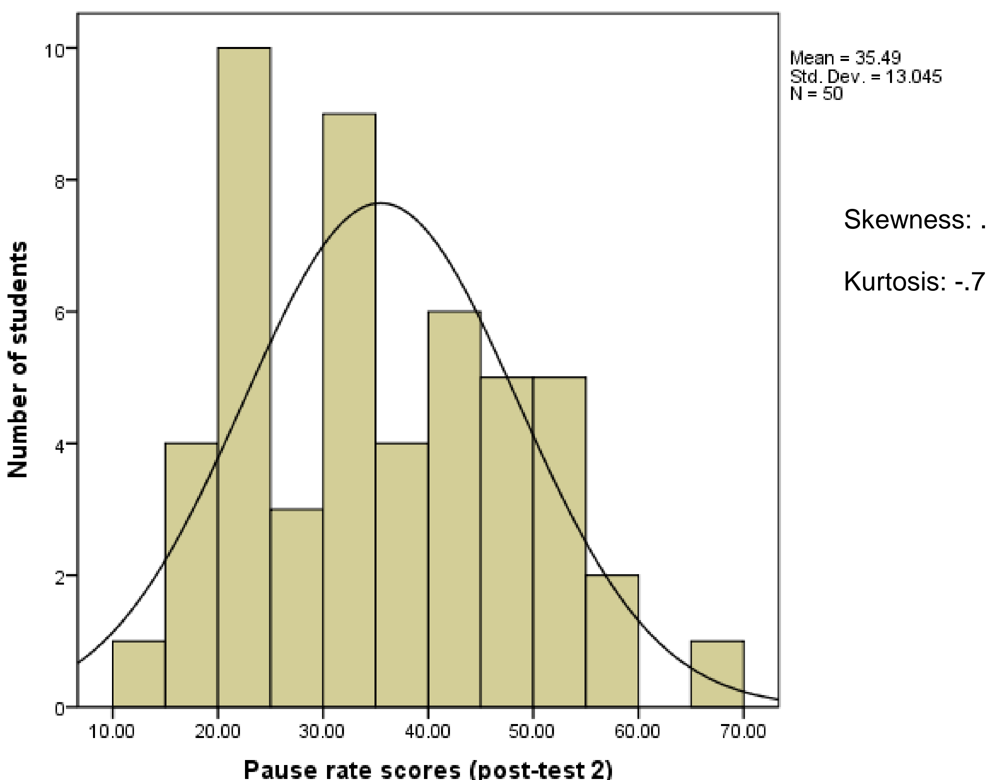
Post-test 1



Skewness: 1.15

Kurtosis: 1.11

Post-test 2



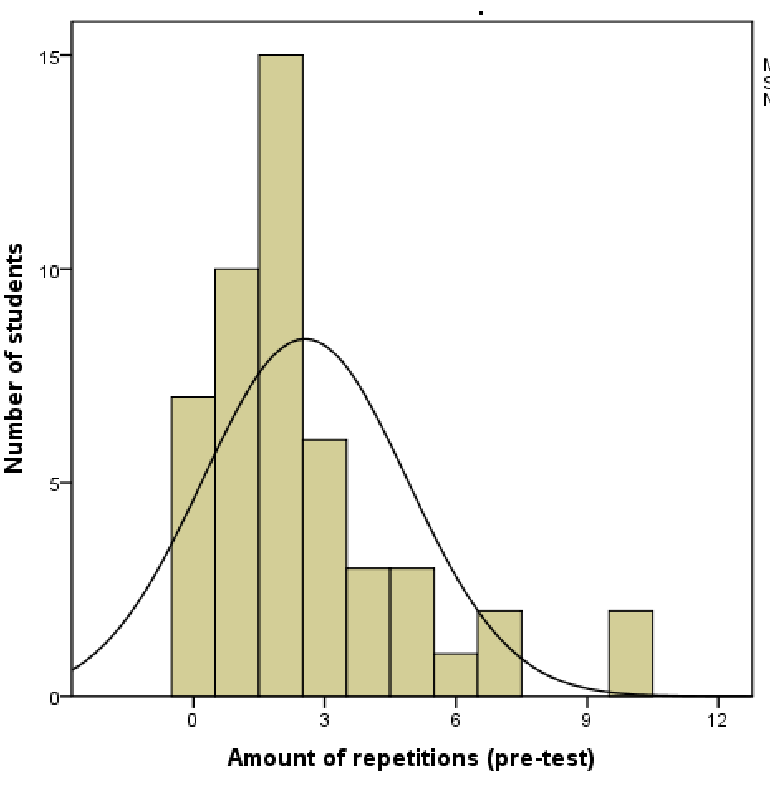
Skewness: .36

Kurtosis: -.72

Appendix 45: Distribution graphs for dysfluency variables of fluency (for proficiency groups)

Repetitions

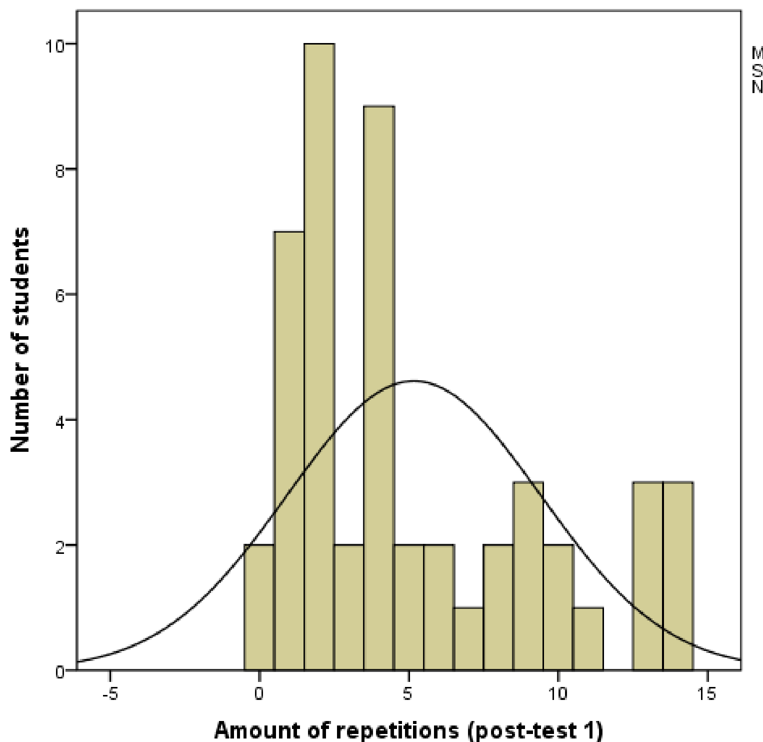
Pre-test



Skewness: 1.60

Kurtosis: 2.79

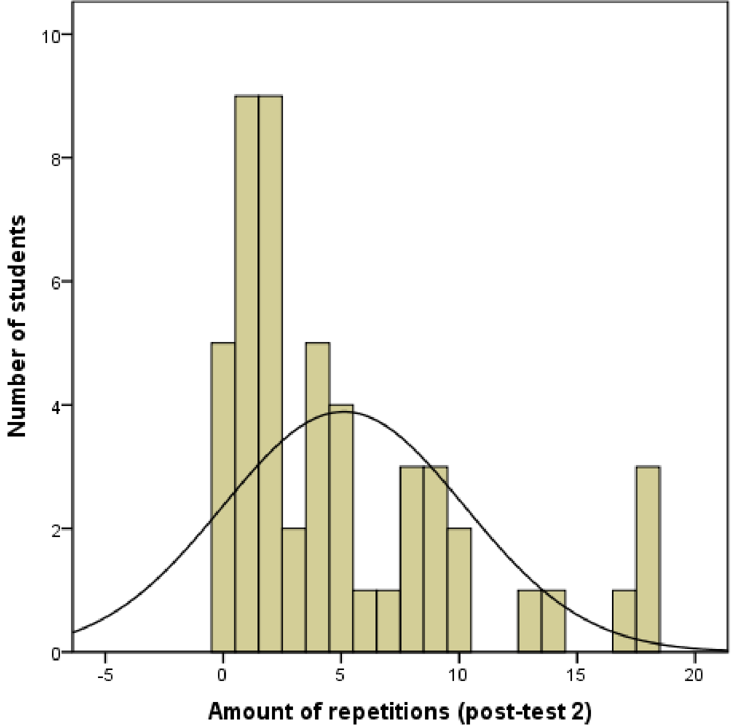
Post-test 1



Skewness: .87

Kurtosis: -.46

Post-test 2

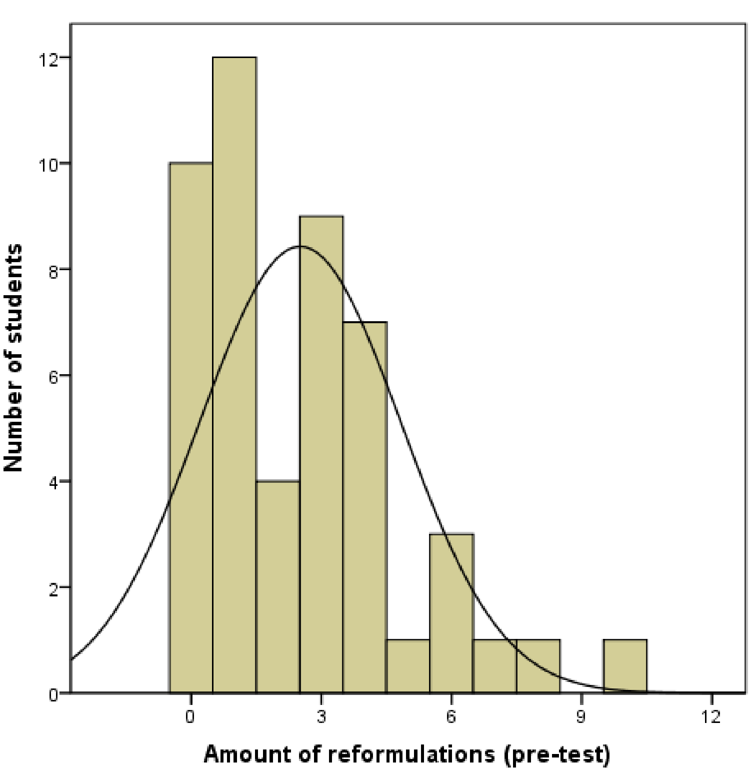


Skewness: 1.29

Kurtosis: .88

<u>Reformulations</u>

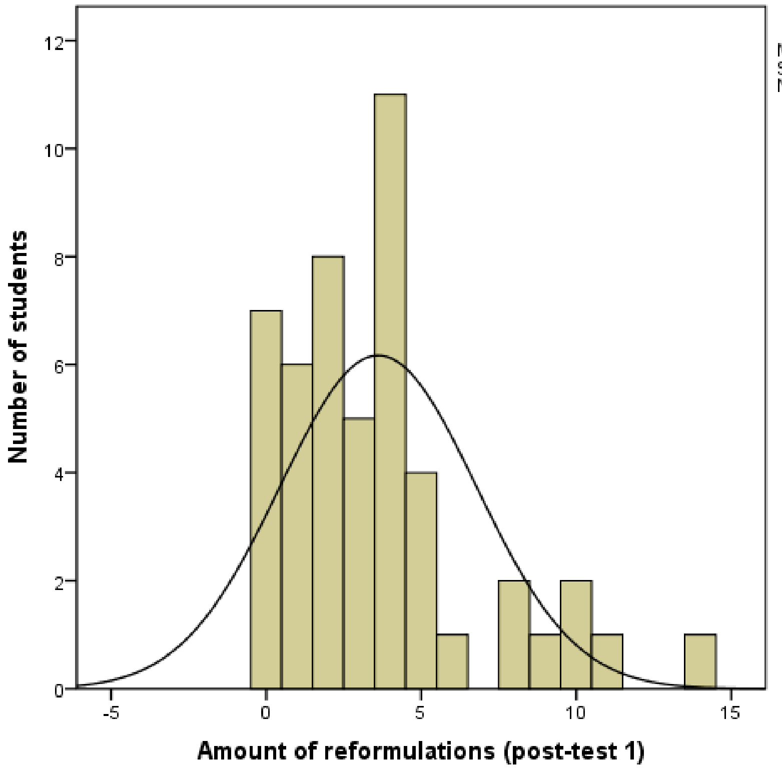
Pre-test



Skewness: 1.12

Kurtosis: 1.24

Post-test 1

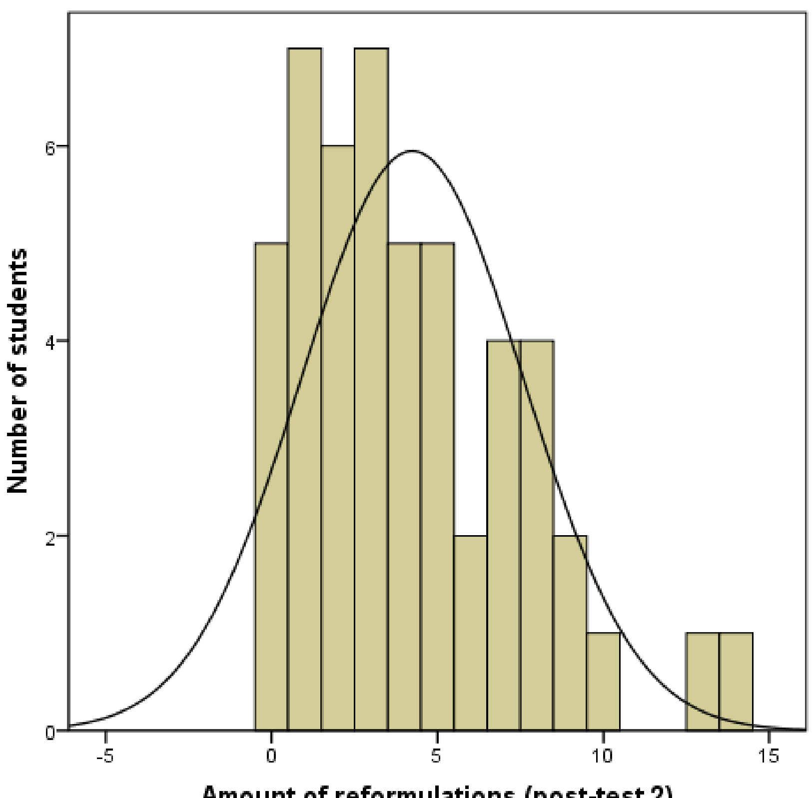


Mean = 3.61
Std. Dev. = 3.168
N = 49

Skewness: 1.31

Kurtosis: 1.75

Post-test 2



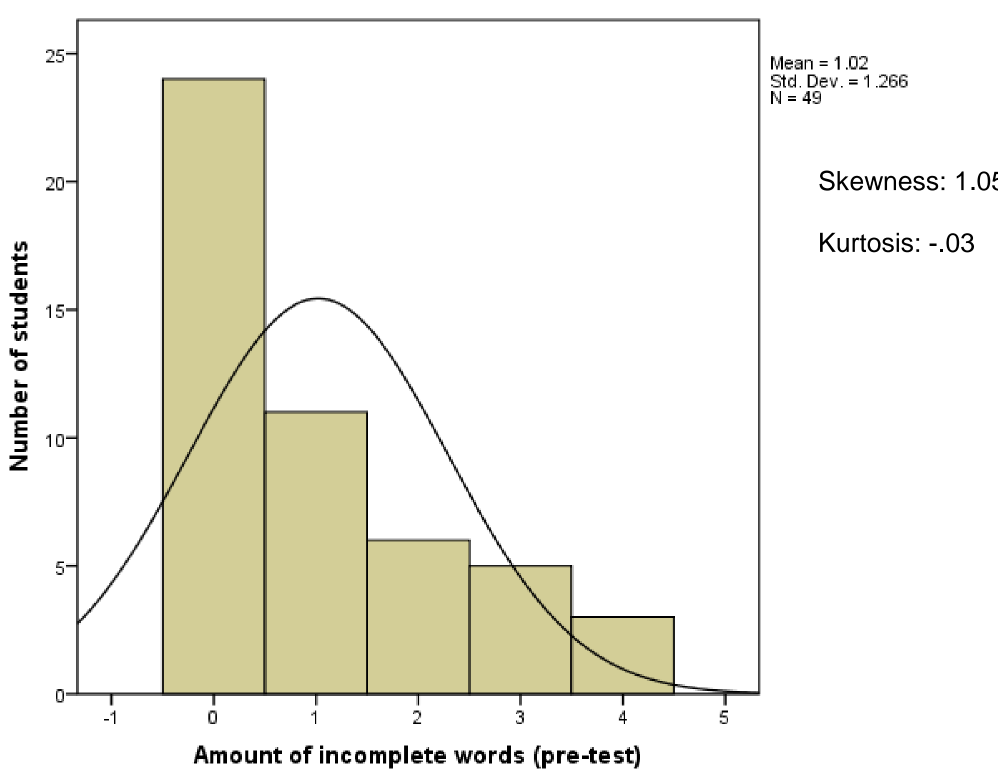
Mean = 4.24
Std. Dev. = 3.354
N = 50

Skewness: .91

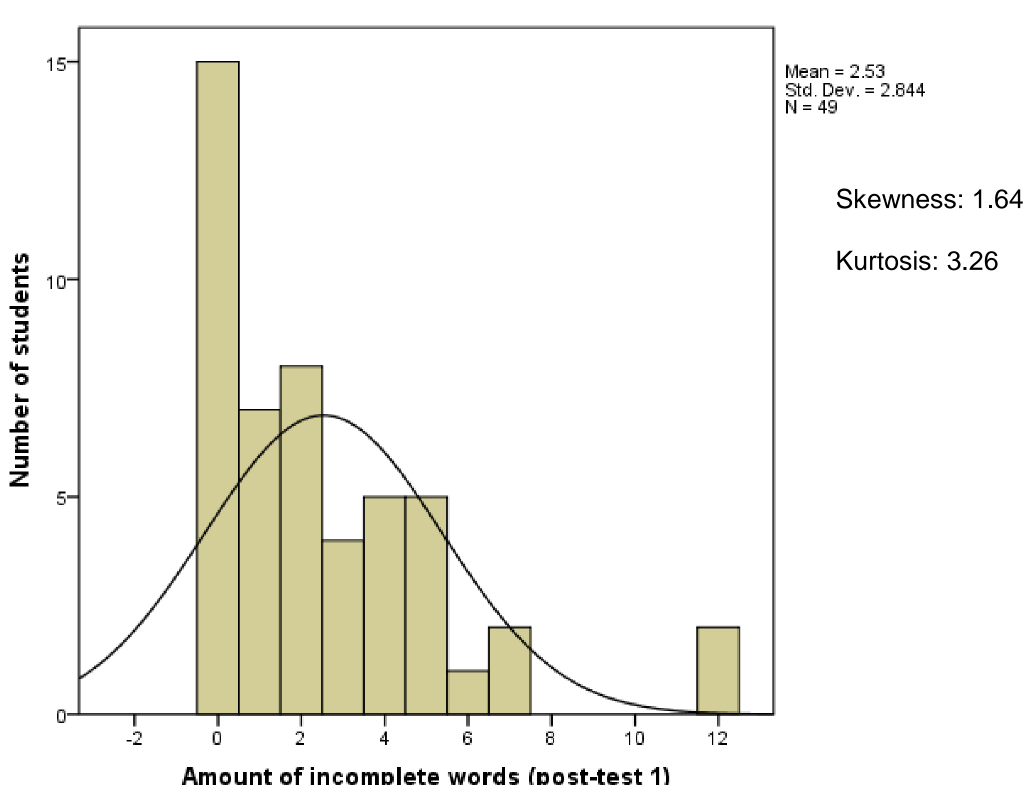
Kurtosis: .58

Incomplete words

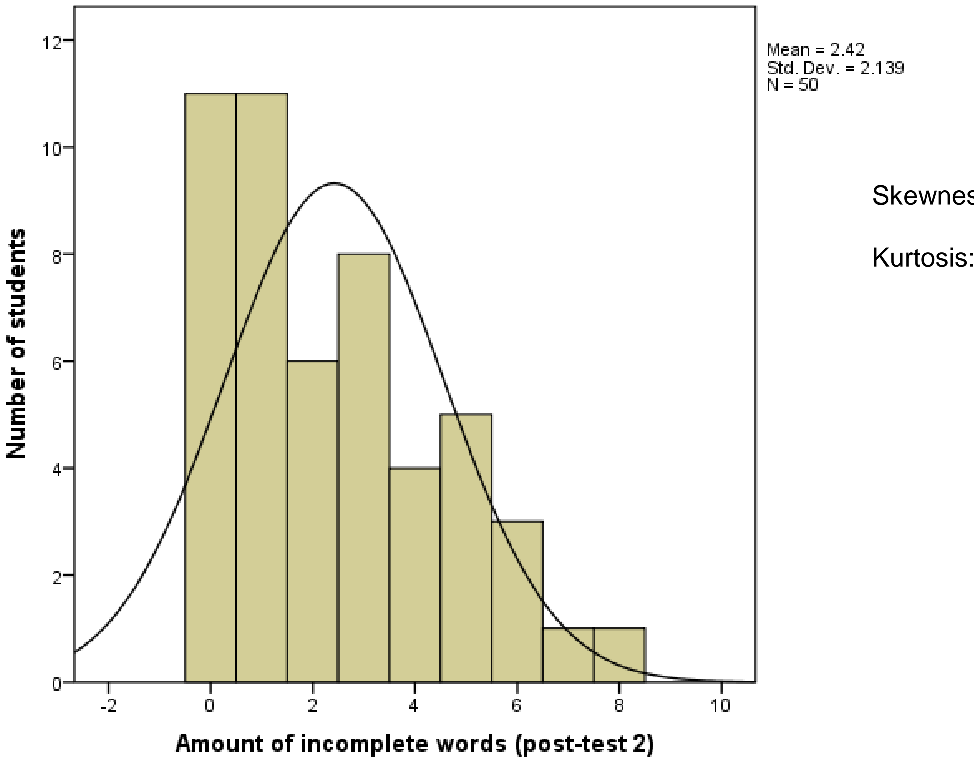
Pre-test



Post-test 1



Post-test 2



Skewness: .70

Kurtosis: -.36

Appendix 46: Comparisons of results of parametric to non-parametric tests for fluency measures (for proficiency groups).

Mean Length of Runs	
Mann-Whitney	**T-test**
Pre-test: $U = 137$, $z = -3.26$, $p = .001$	**Pre-test:** $t\,(33.10) = -3.36$, $p < .01$
Post-test 1: $U = 112$, $z = -3.76$, $p < .001$	**Post-test 1:** $t\,(41.59) = -6.23$, $p < .001$
Post-test 2: $U = 145.5$, $z = -3.24$, $p < .001$	**Post-test 2:** $t\,(48) = -3.36$, $p = .001$

Pause Rate	
Mann-Whitney	**T-test**
Pre-test: $U = 105$, $z = -3.90$, $p < .001$	**Pre-test:** $t\,(47) = 3.85$, $p < .001$
Post-test 1: $U = 100$, $z = -4.00$, $p < .001$	**Post-test 1:** $t\,(37.46) = 6.25$, $p < .001$
Post-test 2: $U = 95$, $z = -6.22$, $p < .001$	**Post-test 2:** $t\,(48) = 5.17$, $p < .001$

Repetitions	
Mann-Whitney	**T-test**
Pre-test: $U = 274$, $z = -.53$, $p > .05$	**Pre-test:** $t\,(38.50) = -1.27$, $p > .05$
Post-test 1: $U = 246$, $z = -1.09$, $p > .05$	**Post-test 1:** $t\,(47) = .81$, $p > .05$
Post-test 2: $U = 266$, $z = -.91$, $p > .05$	**Post-test 2:** $t\,(48) = -.03$, $p > .05$

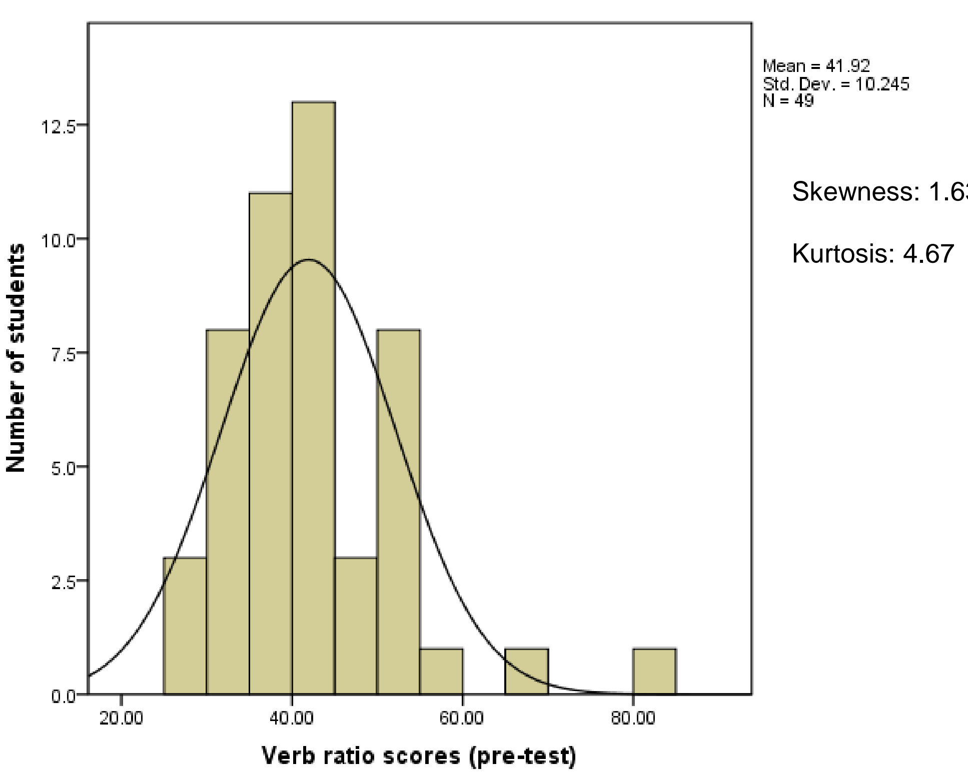
Reformulations	
Mann-Whitney	**T-test**
Pre-test: $U = 220.5$, $z = -1.62$, $p > .05$	**Pre-test:** $t\,(47) = -2.07$, $p < .05$
Post-test 1: $U = 285.5$, $z = -.29$, $p > .05$	**Post-test 1:** $t\,(39.13) = -.79$, $p > .05$
Post-test 2: $U = 285.5$, $z = -.51$, $p > .05$	**Post-test 2:** $t\,(48) = -.59$, $p > .05$

Incomplete words	
Mann-Whitney	**T-test**
Pre-test: $U = 269.50$, $z = -.65$, $p > .05$	**Pre-test:** $t\,(47) = .34$, $p > .05$
Post-test 1: $U = 268.50$, $z = -.64$, $p > .05$	**Post-test 1:** $t\,(47) = -.07$, $p > .05$
Post-test 2: $U = 248.50$, $z = -1.26$, $p > .05$	**Post-test 2:** $t\,(48) = 1.40$, $p > .05$

Appendix 47: Distribution graphs for content variables (for proficiency groups)

Verb Ratio

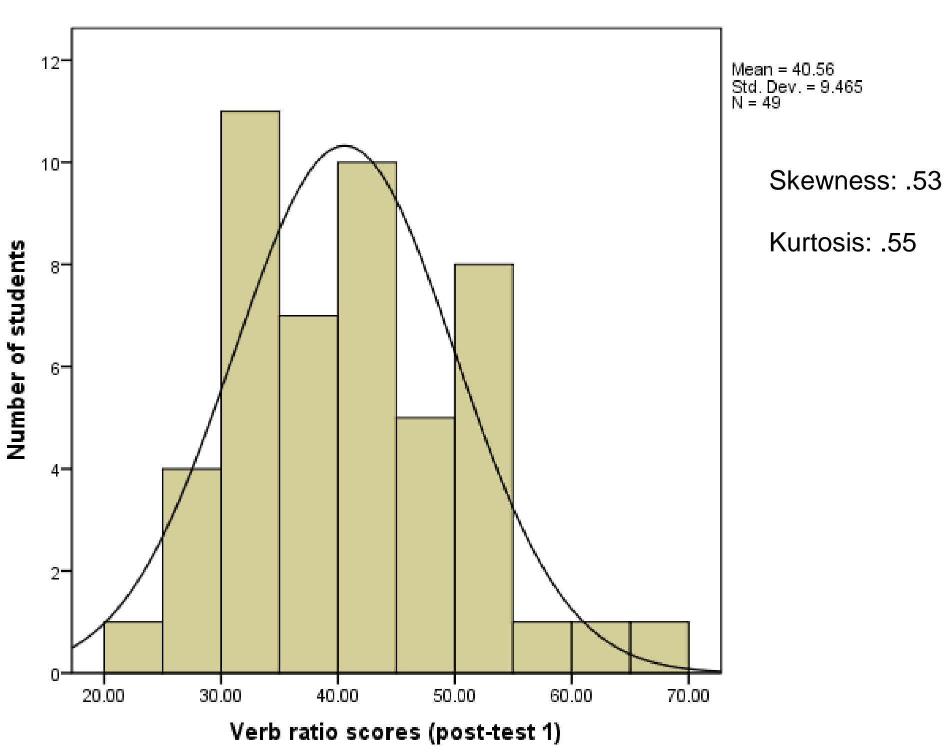
Pre-test



Skewness: 1.63

Kurtosis: 4.67

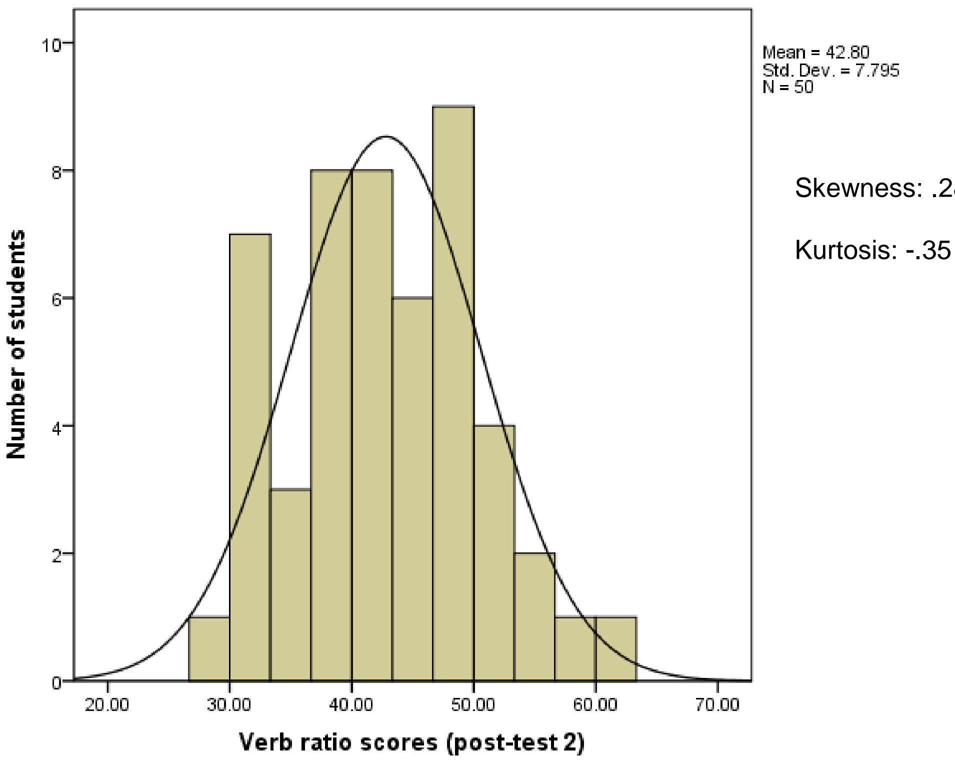
Post-test 1



Skewness: .53

Kurtosis: .55

Post-test 2

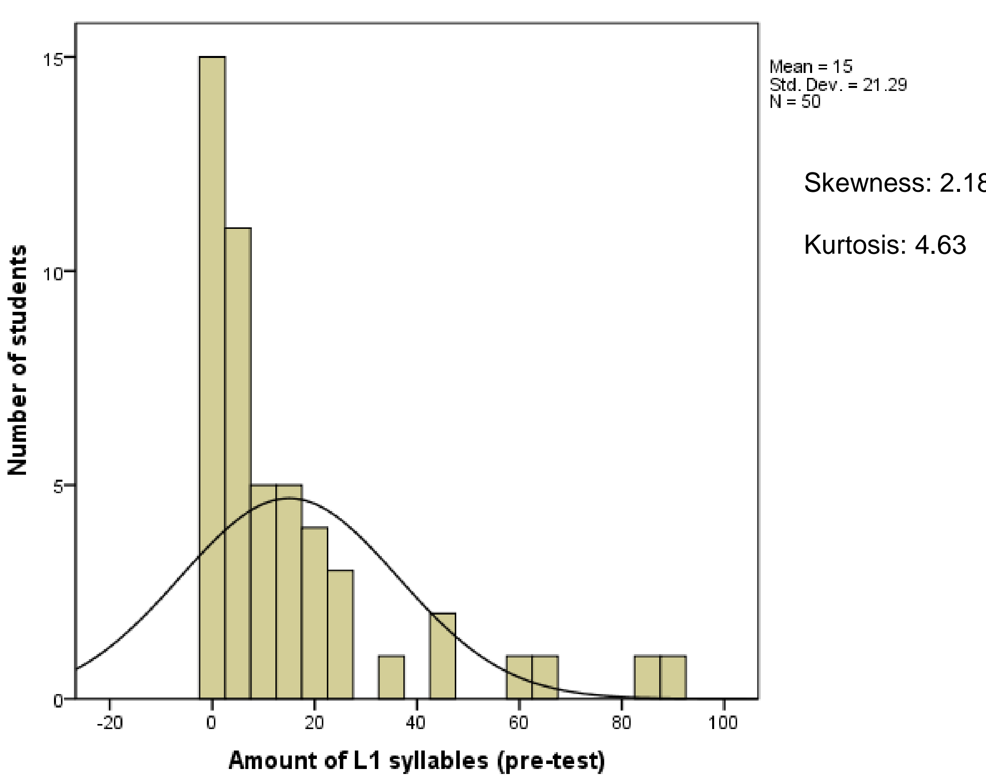


Skewness: .28

Kurtosis: -.35

L1 syllables

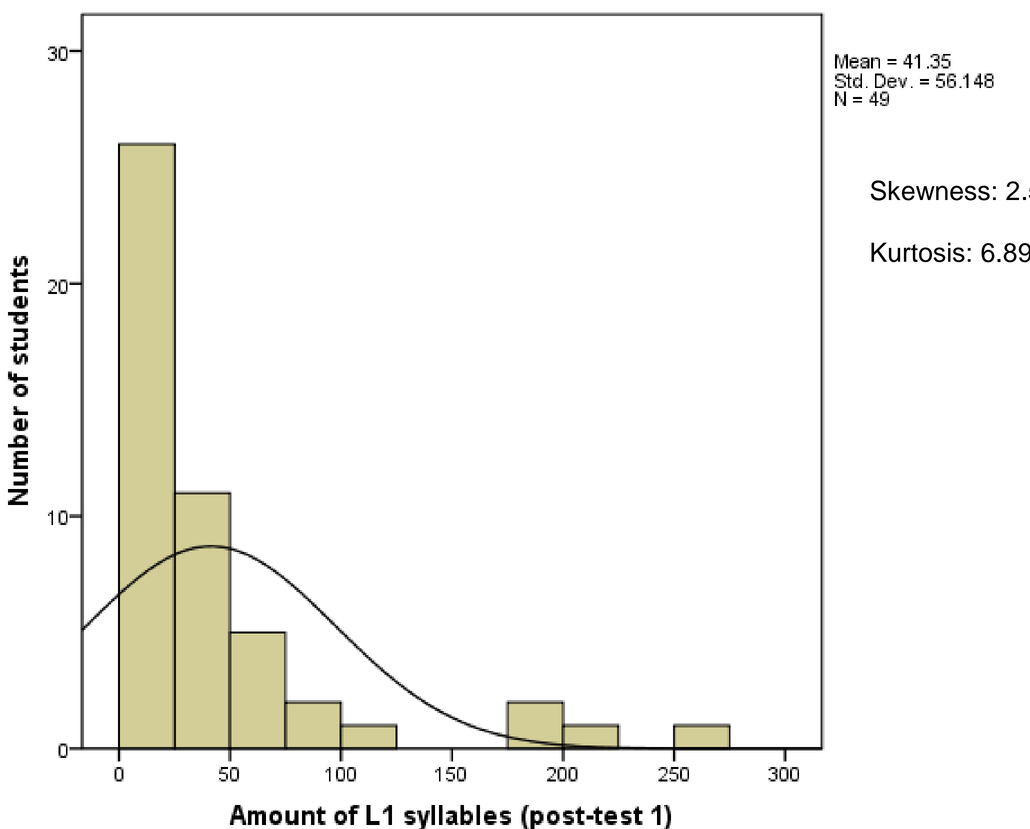
Pre-test



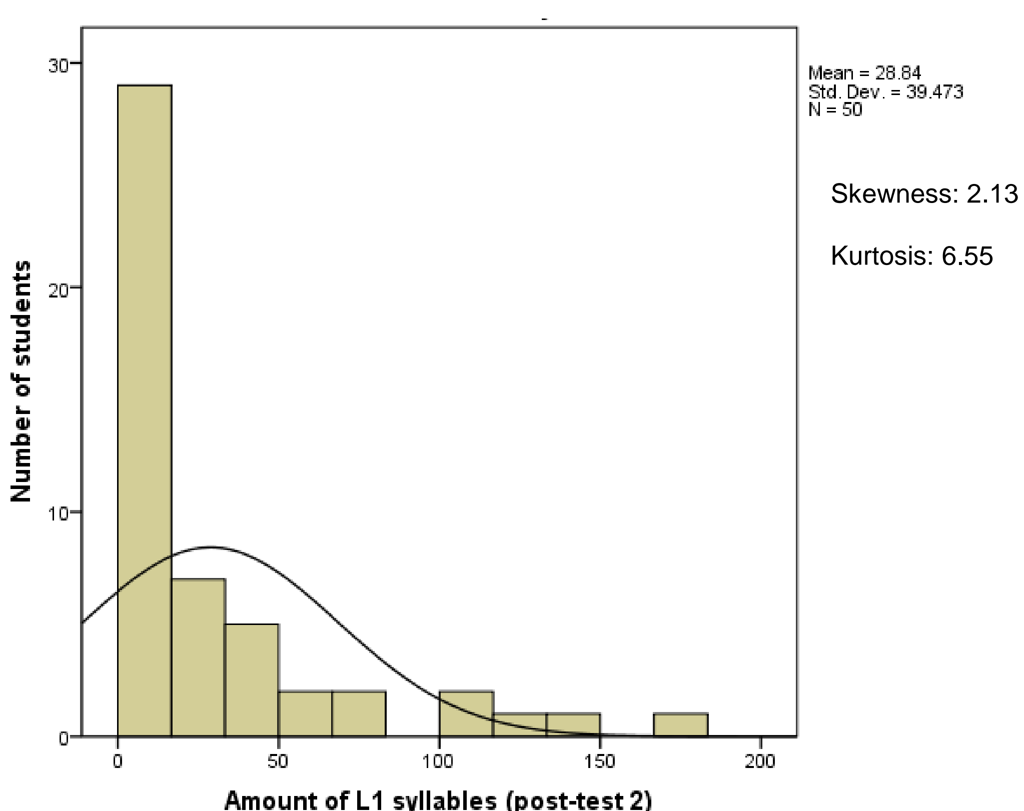
Skewness: 2.18

Kurtosis: 4.63

Post-test 1

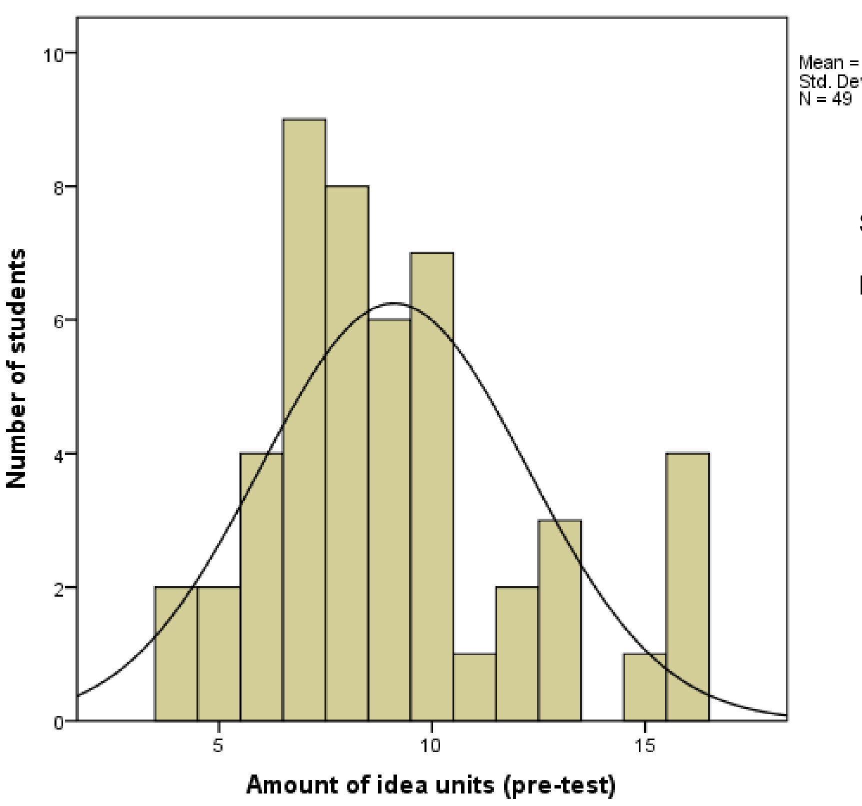


Post-test 2



Idea units

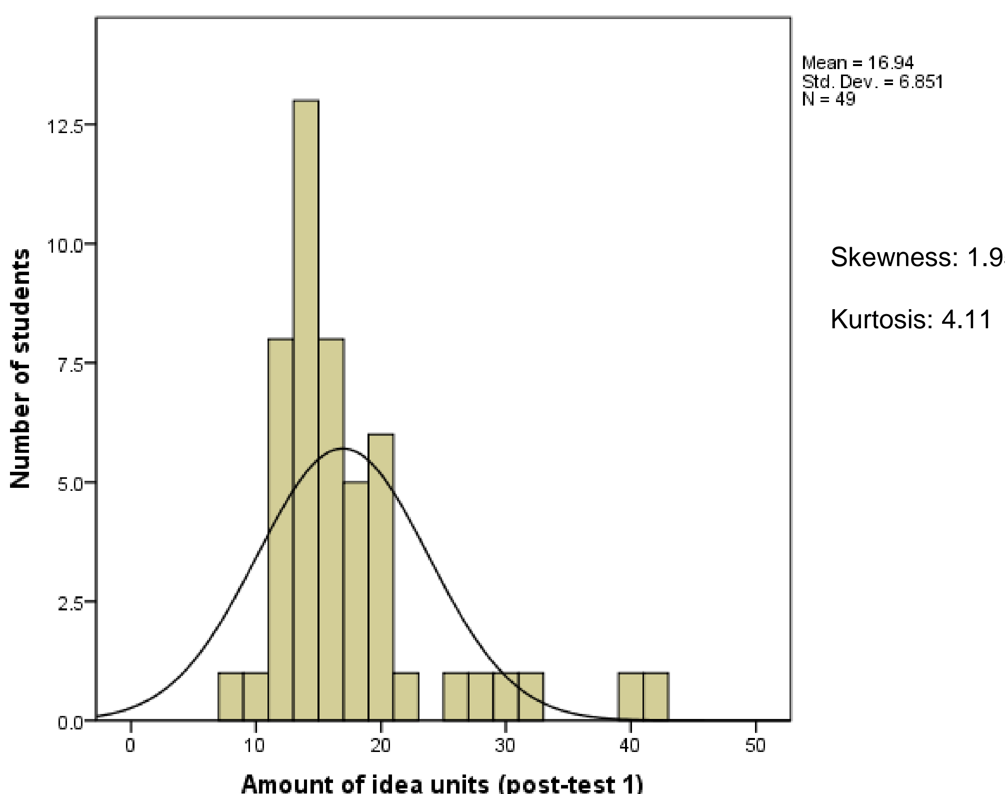
Pre-test



Skewness: .81

Kurtosis: .14

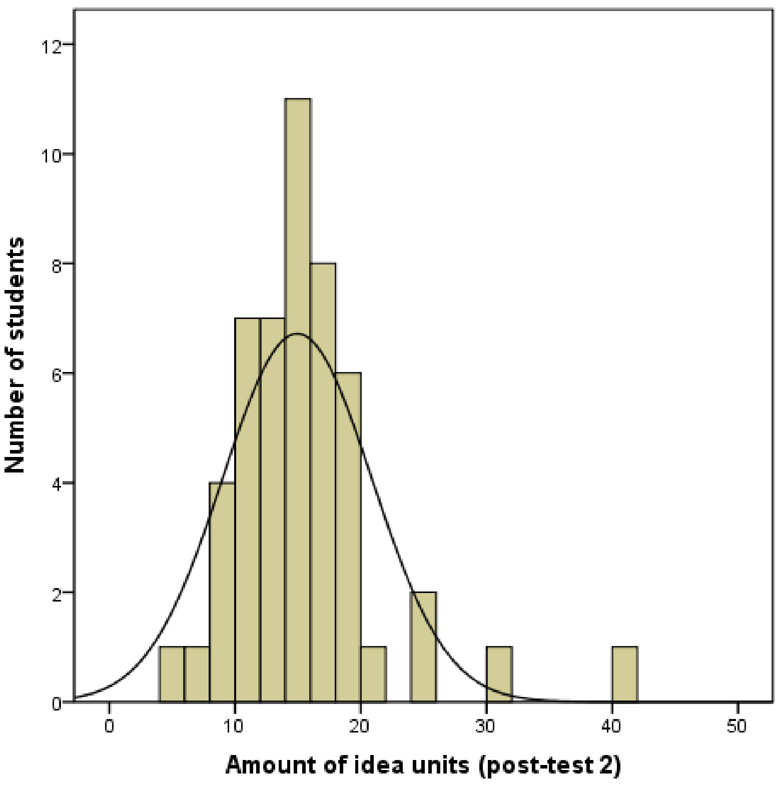
Post-test 1



Skewness: 1.95

Kurtosis: 4.11

Post-test 2



Skewness: 1.91

Kurtosis: 6.27

Appendix 48: Comparisons of results of parametric to non-parametric tests for content measures (for proficiency groups).

Verb Ratio	
Mann-Whitney	**T-test**
Pre-test: $U = 290$, $z = -.20$, $p > .05$	**Pre-test:** $t\,(47) = -.11$, $p > .05$
Post-test 1: $U = 282$, $z = -.36$, $p > .05$	**Post-test 1:** $t\,(47) = .12$, $p > .05$
Post-test 2: $U = 301.5$, $z = -.21$, $p > .05$	**Post-test 2:** $t\,(48) = -3.36$, $p > .05$

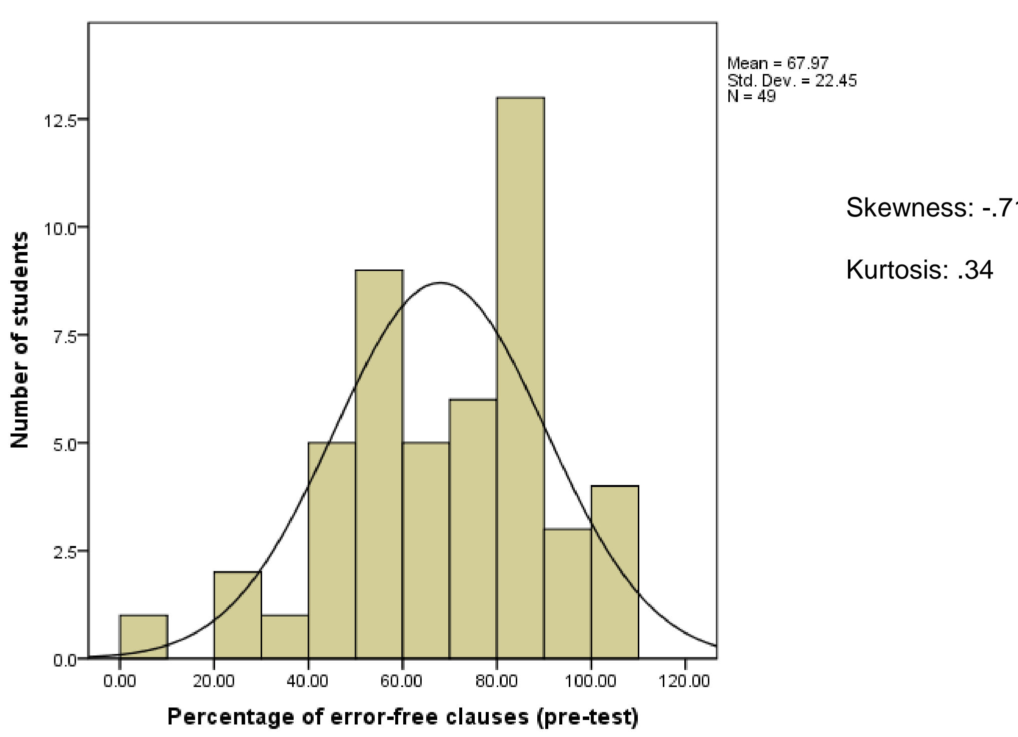
L1 syllables	
Mann-Whitney	**T-test**
Pre-test: $U = 192$, $z = -2.36$, $p < .05$	**Pre-test:** $t\,(48) = 3.15$, $p < .01$
Post-test 1: $U = 140$, $z = -3.20$, $p = .001$	**Post-test 1:** $t\,(26.88) = 2.95$, $p < .01$
Post-test 2: $U = 149.5$, $z = -3.18$, $p = .001$	**Post-test 2:** $t\,(27.16) = 3.57$, $p = .001$

Idea Units	
Mann-Whitney	**T-test**
Pre-test: $U = 156$, $z = -2.90$, $p < .01$	**Pre-test:** $t\,(41.36) = -3.39$, $p < .01$
Post-test 1: $U = 115.50$, $z = -3.71$, $p < .001$	**Post-test 1:** $t\,(29.80) = -3.93$, $p < .001$
Post-test 2: $U = 178$, $z = -2.62$, $p < .01$	**Post-test 2:** $t\,(48) = -2.77$, $p < .01$

Appendix 49: Distribution graphs for accuracy variables (for proficiency groups)

Percentage of error-free clauses

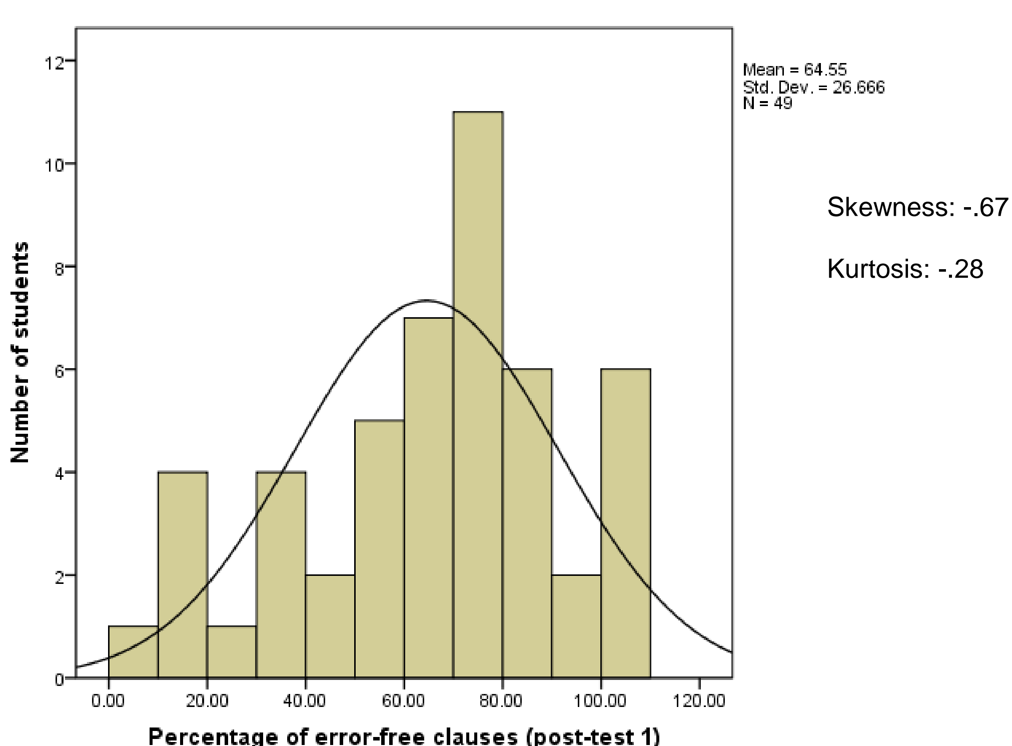
Pre-test



Skewness: -.71

Kurtosis: .34

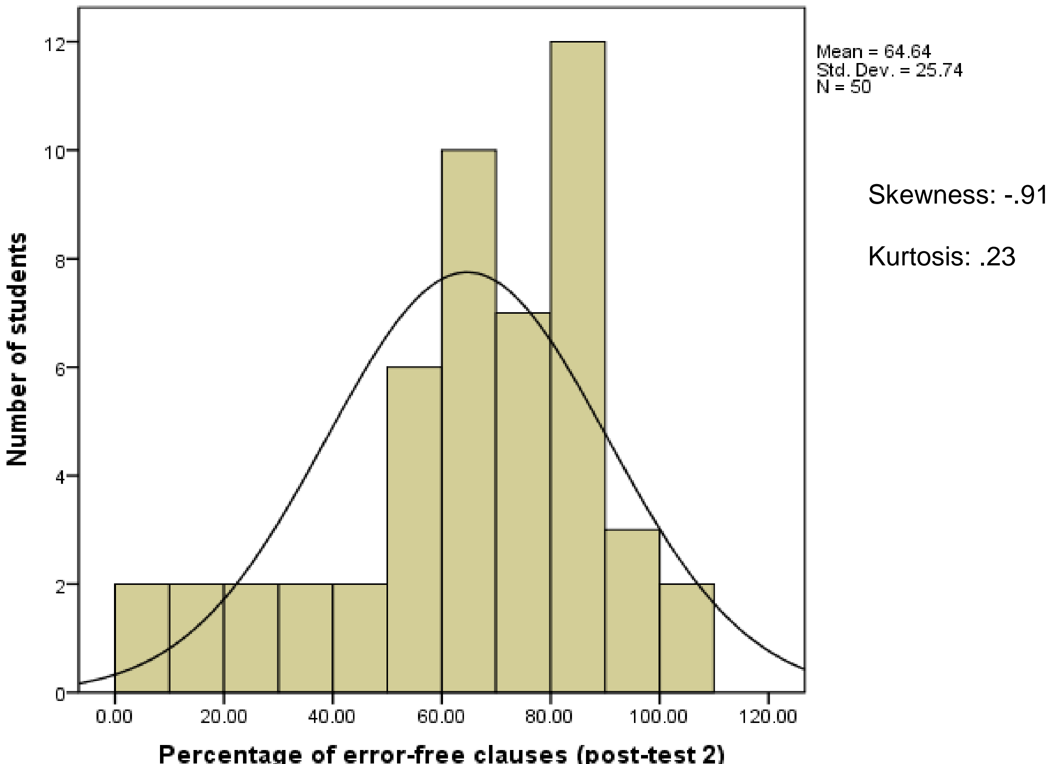
Post-test 1



Skewness: -.67

Kurtosis: -.28

Post-test 2

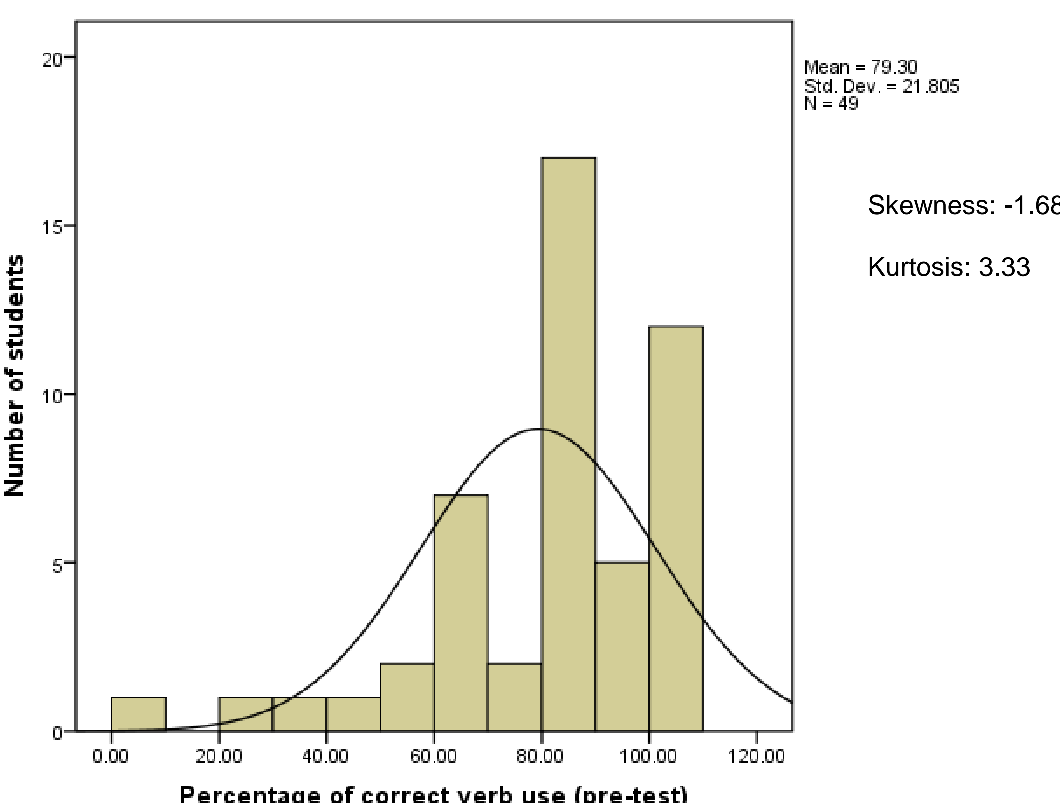


Skewness: -.91

Kurtosis: .23

Percentage of correct verb use

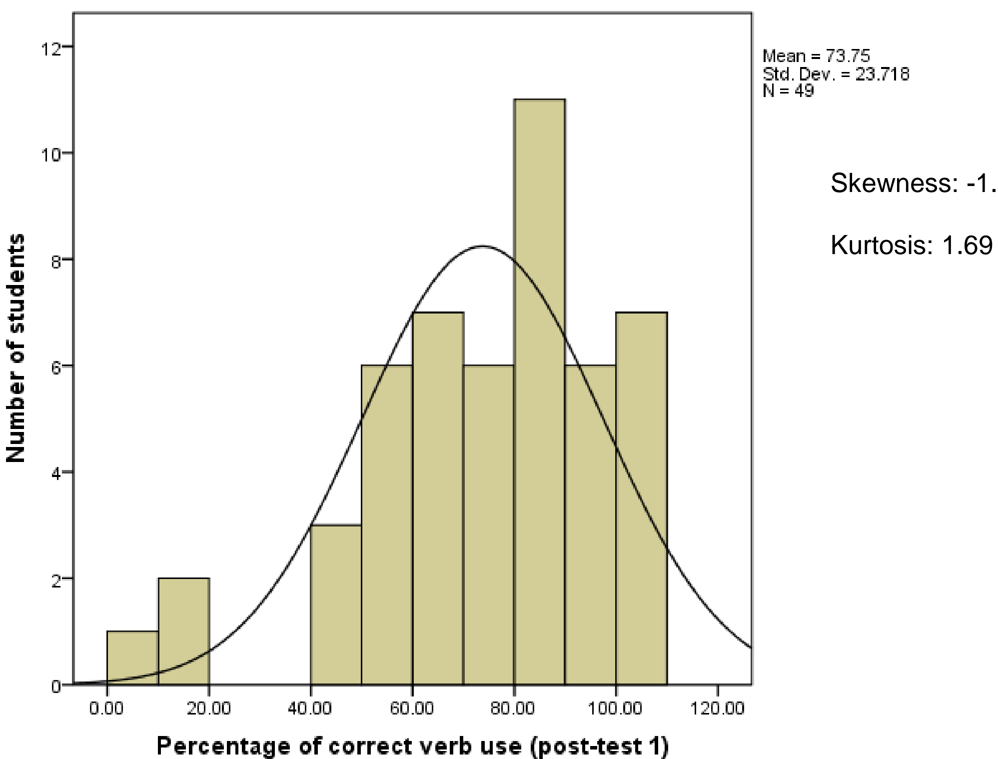
Pre-test



Skewness: -1.68

Kurtosis: 3.33

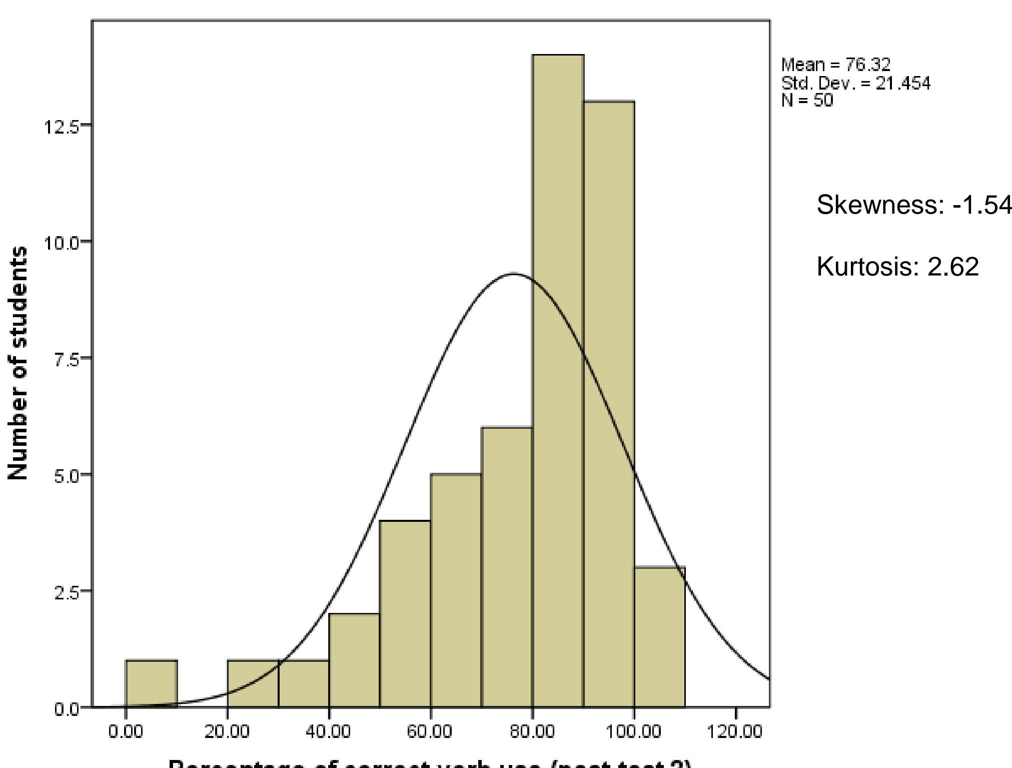
Post-test 1



Skewness: -1.27

Kurtosis: 1.69

Post-test 2



Skewness: -1.54

Kurtosis: 2.62

Appendix 50: Comparisons of results of parametric to non-parametric tests for accuracy measures (for proficiency groups).

Percentage of error-free clauses	
Mann-Whitney	**T-test**
Pre-test: $U = 143$, $z = -3.14$, $p < .01$	**Pre-test:** $t\,(47) = -3.45$, $p = .001$
Post-test 1: $U = 118.5$, $z = -3.63$, $p < .001$	**Post-test 1:** $t\,(38.22) = -6.29$, $p < .001$
Post-test 2: $U = 128.5$, $z = -3.57$, $p < .001$	**Post-test 2:** $t\,(48) = -3.82$, $p < .001$

Percentage of correct verb use	
Mann-Whitney	**T-test**
Pre-test: $U = 177.50$, $z = -2.47$, $p < .05$	**Pre-test:** $t\,(30.31) = -2.94$, $p < .01$
Post-test 1: $U = 128.5$, $z = -3.44$, $p = .001$	**Post-test 1:** $t\,(34.65) = -3.83$, $p = .001$
Post-test 2: $U = 139.5$, $z = -3.36$, $p = .001$	**Post-test 2:** $t\,(48) = -2.96$, $p < .01$