The effectiveness of explicit grammar instruction for the young foreign language learner: A classroom-based experimental study

Rowena Eloise Hanan

PhD

University of York

Education

March 2015

Abstract

Within the input-poor foreign language classroom, opportunities to draw on implicit learning mechanisms are limited. Yet little research has explored the effectiveness of explicit instruction for young learners. The current study investigated the extent to which English learners of L2 German (aged 9-11) benefitted from instruction on accusative case-marking (*den*) for masculine definite articles in German, a problematic feature for L1 English learners due to a reliance on word order when assigning grammatical roles (as predicted by MacWhinney's Competition Model and VanPatten's First Noun Principle).

Two input-based interventions provided explicit information plus EITHER: Task Essential Form-Meaning Connection (TE-FMC) activities forcing attention on the article *and* its role-assigning function; OR Task Essential-Form (TE-F) activities forcing attention on the article only ('spot the form'). Learners were randomly assigned to the TE-FMC (*n* = 45) and TE-F (*n* = 41) treatments. A control group (*n* = 52) received instruction on lexical items, but no exposure to *den*. Two untimed written tasks (sentence matching, gap fill), three one-to-one oral tasks (act-out comprehension, act-out production, elicited imitation), and a metalinguistic task were administered as pre-, post-, and delayed post-tests to assess knowledge of *der* and *den*.

Both interventions yielded large, durable gains across the written and oral tasks. The Control group made no improvement. The TE-FMC and TE-F learners' verbalisable knowledge also improved at post-test, but deteriorated by delayed post-test. Under both conditions, learners had developed explicit knowledge of the target feature, available on untimed written tasks, as well as more automatized knowledge, accessible under time and communicative pressure. Fine-grained analysis revealed that group-level gains could be accounted for by a sub-group of learners within each condition, reflecting the influence of individual differences on instructional effectiveness. The findings contribute to previous research by demonstrating the beneficial role of explicit instruction and knowledge for child L2 learning.

Table of Contents

Abstract 2
Table of Contents 3
List of tables 11
List of figures 14
Acknowledgements 16
Author's declaration 17

CHAPTER 1: INTRODUCTION 18
1.1 The research context 18
1.2 The educational context: Overview of the state education system in England 19
1.3 Changes to foreign language policy in England 20
1.4 Foreign language teaching at KS2 21
1.5 Foreign language grammar teaching at KS2 22
  1.5.1 Curriculum guidelines and current practice 22
  1.5.2 Promoting language awareness 23
1.6 Outline of the thesis 26
CHAPTER 2: LITERATURE REVIEW 27
2.1 How do children acquire a second language? 27
  2.1.1 Age effects in SLA 27
  2.1.2 Critical Period Hypothesis 27
    2.1.2.1 Theories of child language acquisition 29
    2.1.2.2 Does younger equal better? 31
  2.1.3 Child language acquisition and explicit grammar instruction 32
    2.1.3.1 …in the L1 32
    2.1.3.2 …in the L2 34
2.2 Role of explicit knowledge in language acquisition 39
  2.2.1 Learning-acquisition / explicit-implicit dichotomy 39
  2.2.2 Explicit versus implicit knowledge 39
  2.2.3 Non-interface hypothesis 39
  2.2.4 Strong interface hypothesis 41
  2.2.5 Weak interface hypothesis 42
  2.2.6 Operationalising explicit and implicit knowledge 43

2.3 Explicit grammar instruction and SLA 44
2.3.1 Focus on form, focus on formS, or form-focussed? 45
2.3.2 Role of input 47
2.3.2.1 Noticing, attention, and awareness in SLA 48
2.3.2.2 Barriers to noticing 52
2.3.2.3 Consciousness-raising tasks 56
2.3.2.4 Enriched versus enhanced input 57
2.3.3 Input Processing theory 59
2.3.3.1 Principles of Input Processing theory 61
2.3.3.2 Tenets of Principle 2 62
2.3.3.3 Subprinciples of Principle 2 64
2.3.3.4 Challenges to principle 2: Alternative explanations 66
2.3.3.5 Summary of principle 2 68
2.3.4 Processing Instruction 68
2.3.4.1 Core components of PI 69
2.3.4.2 Investigating the effectiveness of processing instruction 73
2.3.4.3 Structured versus enhanced input 77
2.3.4.4 Role of explicit information 80
2.3.4.5 Processing instruction for young learners 82
2.3.4.6 Does PI develop implicit (-awareness) as well as explicit (+awareness) knowledge? 83
2.3.5 Methodological issues in effect of instruction research 84
2.4 Grammatical sensitivity and language learning 86
2.4.1 Defining grammatical sensitivity 86
2.4.2 Grammatical sensitivity and instruction 87
2.4.3 Grammatical sensitivity and young learners 88
2.5 Target grammatical feature 89
2.6 Rationale and research questions 91
CHAPTER 3: METHODOLOGY AND METHODS 93
3.1 Classroom-based experimental research 93
3.1.1 Classroom- versus laboratory-based experimental studies 93
3.1.2 Experimental research 94
3.1.2.1 Characteristics of a formal experiment 94
3.1.2.2 Threats to validity 94
3.1.3 Ethical considerations and how they were addressed in the present study 96

3.2 Selecting schools ... 98
3.3 Participants ... 98
3.3.1 Matched pair randomisation ... 99
3.3.2 Control group ... 100
3.3.3 Pre-teaching phase ... 101
3.4 Design of the study ... 102
3.4.1 Experimental procedure ... 102
3.4.2 Piloting the intervention and test materials ... 103
3.4.2.1 Pilot study procedure ... 103
3.4.2.2 Main findings and implications of the pilot study ... 105
3.5 The Intervention ... 106
3.5.1 Designing tasks for young learners ... 106
3.5.2 Designing the intervention materials ... 107
3.5.3 Explicit information ... 108
3.5.4 TE-FMC intervention activities ... 111
3.5.4.1 *Which picture?* activity (TE-FMC) ... 112
3.5.4.2 *Who's doing what?* activity (TE-FMC) ... 115
3.5.4.3 *Missing nouns* activity (TE-FMC) ... 116
3.5.5 TE-F intervention activities ... 117
3.5.5.1 *Picture Matching* task (TE-F) ... 119
3.5.5.2 Noticing task (TE-F) ... 121
3.5.5.3 *Sensible or silly?* task (TE-F) ... 123
3.5.5.4 *Missing noun* activity (TE-F) ... 124
3.5.6 Administering the intervention ... 125
3.6 Outcome measures ... 127
3.6.1 Measuring different types of knowledge ... 127
3.6.2 Controlling animacy ... 129
3.6.3 Written outcome measures ... 130
3.6.3.1 Designing the written activities ... 130
3.6.3.2 *Sentence Matching* task ... 130
3.6.3.3 *Gap-fill* task ... 132
3.6.4 Oral outcome measures ... 133
3.6.4.1 *Act-Out Comprehension* task ... 133
3.6.4.2 *Act-Out Production* task ... 135
3.6.4.3 *Sentence Repetition* task ... 136

3.6.5 *Sentence Reconstruction* task 137
3.6.6 Vocabulary test 140
3.6.7 Grammatical sensitivity test 141
3.6.8 Administering the outcome measures........ 143
3.6.9 Missing data 144
3.7 Statistical analysis 145
3.7.1 Parametric versus Non-parametric........ 145
3.7.1.1 Normality of distribution 146
3.7.1.2 Homogeneity of variance........ 147
3.7.2 Non-parametric analysis 147
3.7.2.1 Non-parametric tests 147
3.7.2.2 Statistical significance 148
3.7.2.3 Calculating effect size 149
3.7.3 Principal component analysis 150
3.7.3.1 Determining the suitability of data for PCA 151
3.7.3.2 Extracting factors 152
3.7.4 Exploring individual participant performance on the outcome measures: Chi-square and McNemar test 153
3.7.5 Exploring the relationship between the outcome measures and the grammatical sensitivity task........ 154
3.7.5.1 ANCOVA 154
3.7.5.2 Correlation and standard regression 155
CHAPTER 4: RESULTS 1 156
4.1 Normality and homogeneity of the dataset 157
4.2 Establishing the baseline........ 158
4.2.1 Equivalence of pre-test performance (by age and school)........ 158
4.2.2 Performance on the vocabulary test 158
4.3 Performance on the written outcome measures 160
4.3.1 *Sentence Matching* task 160
4.3.1.1 Analysis of total scores........ 160
4.3.1.2 Analysis by test condition (SVO, OVS)........ 161
4.3.1.3 Examination of effect sizes 164
4.3.2 *Gap-fill* task 164
4.3.2.1 Analysis of total scores........ 164
4.3.2.2 Analysis by test condition (SVO+Subj, OVS+Subj, SVO+Obj, OVS+Obj) 166
4.3.2.3 Examination of effect sizes 171

4.3.3 Summary of the findings for the written outcome measures 171
4.4 Performance on the oral outcome measures 172
4.4.1 *Act-Out Comprehension* task 172
4.4.1.1 Analysis of total scores.. 172
4.4.1.2 Analysis by test condition (SVO, OVS)... 174
4.4.1.3 Examination of effect sizes 176
4.4.2 *Act-Out Production* task 176
4.4.2.1 Analysis of total scores.. 176
4.4.2.2 Analysis by test condition (Subject, Object) 178
4.4.2.3 Examination of effect sizes 181
4.4.2.4 Analysis of article use.. 181
4.4.3 *Sentence Repetition* task 184
4.4.3.1 Analysis of total scores.. 184
4.4.3.2 Analysis by test condition (SVO+Subj, OVS+Subj, SVO+Obj, OVS+Obj) 186
4.4.3.3 Examination of effect sizes 190
4.4.3.4 Analysis of article use.. 190
4.4.4 Summary of findings for the oral outcome measures 195
4.5 Analysis of Animacy conditions (A+A, A+I, I+A) 195
4.5.1 Performance on Animacy conditions (A+A, A+I, I+A) at pre- and post-test 196
4.5.1.1 Analysis of total scores.. 196
4.5.1.2 Analysis by test condition (SVO, OVS)... 197
4.5.2 Gains made on the Animacy conditions (A+A, A+I, I+A) 199
4.5.3 Summary of findings for the Animacy conditions... 200
4.6 Performance on the metalinguistic task 200
4.6.1 *Sentence Reconstruction* task: Quantitative analysis 200
4.6.1.1 Analysis of scores for Order 200
4.6.1.2 Analysis of scores for Explanation 202
4.6.1.3 Examination of effect sizes 203
4.6.1.4 Summary of findings for the quantitative analysis 203
4.6.2 *Sentence Reconstruction* task: Qualitative analysis 204
4.6.2.1 Metalinguistic knowledge at pre-test 204
4.6.2.2 Metalinguistic knowledge at post-test 206
4.6.2.3 Metalinguistic knowledge at delayed post-test.. 208
4.6.2.4 Summary of findings for the qualitative analysis... 211
4.7 Performance on the six outcome measures 211

4.7.1 Overall magnitude of instructional effect ... 211
4.7.2 Summary of main findings ... 212
CHAPTER 5: RESULTS 2 ... 214
5.1 Relationship between the outcome measures: Principal Component Analysis ... 214
5.1.1 Preparing the data for PCA ... 214
5.1.2 PCA of full dataset (All learners) ... 215
5.1.3 PCA of group datasets (TE-FMC, TE-F) ... 217
5.1.4 Summary of PCA findings ... 220
5.2 Individual participant performance on the outcome measures ... 220
5.2.1 Distribution of scores ... 220
5.2.1.1 Examination of histograms ... 220
5.2.1.2 Identification of sub-groups ... 223
5.2.2 *Sentence Matching* task ... 224
5.2.3 *Gap-Fill* task ... 226
5.2.4 *Act-Out Comprehension* task ... 228
5.2.5 *Act-Out Production* task ... 229
5.2.6 *Sentence Repetition* task ... 231
5.2.7 *Sentence Reconstruction (Explanation)* task ... 233
5.2.8 Summary of findings for sub-group analysis ... 235
5.3 Role of grammatical sensitivity ... 236
5.3.1 Performance on the grammatical sensitivity task ... 236
5.3.2 Grammatical sensitivity as a covariate ... 236
5.3.3 Relationship between grammatical sensitivity and performance on the outcome measures ... 238
5.3.3.1 Correlation analysis ... 238
5.3.3.2 Regression analysis ... 239
5.4 Summary of Results 2 ... 241
CHAPTER 6: DISCUSSION ... 243
6.1 Overall performance across the outcome measures ... 243
6.1.1 General findings ... 243
6.1.2 Superiority of the interventions ... 244
6.2 RQ1: Does explicit grammar instruction improve young learners' a) comprehension and b) production of the target grammatical feature? ... 246
6.2.1 Effect of instruction on comprehension ... 246
6.2.1.1 Performance on comprehension tasks ... 246
6.2.1.2 Overreliance on word order ... 246

6.2.1.3 Overcoming an overreliance on word order 248
6.2.2 Effect of instruction on production 249
6.2.2.1 Performance on production tasks 249
6.2.2.2 Use of the target feature *den* 251
6.2.2.3 Overuse of the target feature *den* 252
6.2.3 Written versus oral tasks 254
6.2.4 Summary of discussion for RQ1 255
6.3 RQ2: To what extent does explicit grammar instruction develop different types of knowledge of the target grammatical feature? 256
6.3.1 Evidence of explicit knowledge 257
6.3.2 Evidence of metalinguistic knowledge 257
6.3.3 Evidence of a different type of knowledge 261
6.3.3.1 A more implicit form of knowledge? 261
6.3.3.2 Methodological considerations 263
6.3.3.3 Automatized explicit knowledge? 264
6.3.4 Summary of discussion for RQ2 265
6.4 RQ3 Following explicit information, is intentional practice in attending to the target form-meaning connection more beneficial than intentional practice in attending to the target grammatical form only? 265
6.4.1 Equivalence of the interventions 265
6.4.2 Accounting for the TE-F group's improvement 266
6.4.2.1 Encouraging engagement with the target grammatical form 266
6.4.2.2 Provision of explicit information 267
6.4.3 An exception: Sentence Repetition task 269
6.4.4 Individual learner differences: variation in overall test performance 271
6.4.5 Role of grammatical sensitivity 273
6.4.5.1 Relationship between grammatical sensitivity and written task performance 274
6.4.5.2 Relationship between grammatical sensitivity and oral task performance 275
6.4.5.3 Relationship between grammatical sensitivity and metalinguistic task performance 276
6.4.5.4 Relationship between grammatical sensitivity and the TE-F intervention 277
6.4.6 Summary of discussion for RQ3 278
CHAPTER 7: CONCLUSION 279
7.1 Summary of the study 279
7.2 Summary of the findings 280
7.3 Limitations and future research 282

7.4 Implications for the primary foreign language classroom 284
7.5 Contributions of the study 287

Appendix 1 Principle 1 of VanPatten's Input Processing theory 289
Appendix 2 Consent form for head teacher / teacher.. 290
Appendix 3 Letter to parents 292
Appendix 4 List of core vocabulary 293
Appendix 5 *Which picture?* activity (TE-FMC)... 294
Appendix 6 *Who's doing what?* activity (TE-FMC)... 296
Appendix 7 *Picture matching* activity (TE-F) 298
Appendix 8 *Sensible or silly?* activity (TE-F) 300
Appendix 9 Control group vocabulary practice 302
Appendix 10 Vocabulary used in the oral tasks 303
Appendix 11 Number of OVS sentences produced on the Act-Out Production task 303
Appendix 12 Coding scheme for Sentence Reconstruction task 304
Appendix 13 Instructions for second marker (Sentence Reconstruction task) 305
Appendix 14 Protocol for oral tasks (set A) 306
Appendix 15 Results of the Shapiro-Wilk test of normality... 310
Appendix 16 Distribution of scores on each task at pre-test (by experimental group) 313
Appendix 17 Distribution of scores on each task at pre-test (by age group) 314
Appendix 18 Distribution of scores on each task at pre-test (by school)................................ 315
Appendix 19 Distribution of scores on each task at post-test (Control group only) 316
Appendix 20 Distribution of scores on each task at delayed post-test 317
Appendix 21 Results of Levene's test of homogeneity of variance 318
Appendix 22 Participant performance by age group.. 319
Appendix 23 Participant performance by school.. 320
Appendix 24 Descriptive statistics for Animacy conditions 321
Appendix 25 Post-test scores for Not Got It sub-group on Sentence Reconstruction (Explanation) task 322
Appendix 26 Descriptive statistics for TE-FMC intervention (week 1) 322
Definitions... 323
References 325

List of tables

Table 2.1: Definite article case-marking in German ... 89

Table 3.1: Time schedule of study ... 102

Table 3.2: Rotation of outcome measure versions (A, B, C) ... 103

Table 3.3: Pilot study procedure ... 104

Table 3.4: Amount of explicit information provided ... 109

Table 3.5: Number of activities and items in intervention sessions (TE-FMC) ... 112

Table 3.6: Number of items including extra feedback (Which picture? TE-FMC) ... 114

Table 3.7: Number of items including extra feedback (Who's doing what? TE-FMC) ... 116

Table 3.8: Number of items including extra feedback (Missing nouns, TE-FMC) ... 117

Table 3.9: Number of activities and items in intervention sessions (TE-F) ... 118

Table 3.10: Number of items including extra feedback (Picture Matching, TE-F) ... 120

Table 3.11: Number of items including extra feedback (Sensible or silly, TE-F) ... 124

Table 3.12: Number of items including extra feedback (Missing noun, TE-F) ... 125

Table 3.13: Outcome measures ... 128

Table 3.14: Rotation of vocabulary test (A and B) ... 141

Table 3.15: Number of participants included in analysis of tasks ... 145

Table 3.16: Non-parametric tests utilised in the study ... 148

Table 3.17: Tests utilised to analyse categorical sub-group variable ... 153

Table 4.1: Descriptive statistics for the vocabulary test ... 159

Table 4.2: Descriptive statistics for the Sentence Matching task ... 160

Table 4.3: Magnitude of instructional effect on the Sentence Matching task ... 164

Table 4.4: Descriptive statistics for the Gap-fill task ... 165

Table 4.5: Magnitude of instructional effect on the Gap-fill task ... 171

Table 4.6: Descriptive statistics for the Act-Out Comprehension task ... 172

Table 4.7: Magnitude of instructional effect on the Act-Out Comprehension task ... 176

Table 4.8: Descriptive statistics for the Act-Out Production task ... 177

Table 4.9: Magnitude of instructional effect on the Act-Out Production task ... 181

Table 4.10: Percentage article use (Act-Out Production) ... 181

Table 4.11: Incorrect use of den for the subject condition (Act-Out Production) ... 182

Table 4.12: Descriptive statistics for the Sentence Repetition task ... 184

Table 4.13: Magnitude of instructional effect on the Sentence Repetition task ... 190

Table 4.14: Percentage article use (Sentence Repetition) ... 191

Table 4.15: Incorrect use of den for the SVO+Subj and OVS+Subj conditions (Sentence Repetition) 192

Table 4.16: Descriptive statistics for the Animacy conditions (Act-Out Comprehension)........ 196

Table 4.17: Descriptive statistics for the Animacy conditions on the SVO test condition (Act-Out Comprehension)........ 197

Table 4.18: Descriptive statistics for the Animacy conditions on the OVS test condition (Act-Out Comprehension)........ 198

Table 4.19: Gains scores for the Animacy conditions by test condition (Act-Out Comprehension)........ 199

Table 4.20: Descriptive statistics for Sentence Reconstruction task – Order........ 201

Table 4.21: Descriptive statistics for Sentence Reconstruction task - Explanation........ 202

Table 4.22: Magnitude of instructional effect on the Sentence Reconstruction task........ 203

Table 4.23: Overall magnitude of instructional effect........ 211

Table 5.1: Suitability of data for PCA........ 215

Table 5.2: Percentage score for each task (All learners)........ 216

Table 5.3: Correlation matrix for six outcome measures (All learners)........ 216

Table 5.4: Factor loading for each task onto extracted component(s) (All learners)........ 217

Table 5.5: Percentage score for each task (by group)........ 217

Table 5.6: Correlation matrix for six outcome measures (by group)........ 218

Table 5.7: Factor loading for each task onto extracted component(s) (by group)........ 219

Table 5.8 Criteria for allocation to sub-group........ 224

Table 5.9: Distribution of learners within each sub-group (Sentence Matching)........ 224

Table 5.10: Contingency table for Chi-square analysis (Sentence Matching)........ 225

Table 5.11: Crosstabulation of sub-group*time (Sentence Matching)........ 225

Table 5.12: Distribution of learners within each sub-group (Gap-fill)........ 226

Table 5.13: Contingency table for Chi-square analysis (Gap-fill)........ 227

Table 5.14: Crosstabulation of sub-group*time (Gap-fill)........ 227

Table 5.15: Distribution of learners within each sub-group (Act-Out Comprehension)........ 228

Table 5.16: Contingency table for Chi-square analysis (Act-Out Comprehension)........ 228

Table 5.17: Crosstabulation of sub-group*time (Act-Out Comprehension)........ 229

Table 5.18: Distribution of learners within each sub-group (Act-Out Production)........ 230

Table 5.19: Contingency table for Chi-square analysis (Act-Out Production)........ 230

Table 5.20: Crosstabulation of sub-group*time (Act-Out Production)........ 231

Table 5.21: Distribution of learners within each sub-group (Sentence Repetition)........ 231

Table 5.22: Contingency table for Chi-square analysis (Sentence Repetition)........ 232

Table 5.23: Crosstabulation of sub-group*time (Sentence Repetition)........ 232

Table 5.24: Distribution of learners within each sub-group (Sentence Reconstruction-Explanation) 233

Table 5.25: Contingency table for Chi-square analysis (Sentence Reconstruction - Explanation) 234

Table 5.26: Crosstabulation of sub-group*time (Sentence Reconstruction - Explanation) 234

Table 5.27: Descriptive statistics for the grammatical sensitivity task 236

Table 5.28: ANCOVA results for effect of covariate (grammatical sensitivity task score)........ 237

Table 5.29: Results of Pearson's correlation between grammatical sensitivity task scores and outcome measure scores.. 239

Table 5.30: Results of standard regression analysis (predictor variable: grammatical sensitivity task scores) 240

Table 6.1: Post-test scores for two learners from the Not Got It sub-group of Sentence Reconstruction (Explanation) task 259

Table 6.2: Post-test scores on the six outcome measures for two learners from Got It sub-group of Sentence Reconstruction (Explanation) task 259

List of figures

Figure 3.1: Distribution of participants across experimental groups 100
Figure 3.2: Experimental procedure 102
Figure 3.3: Recap of grammatical rule (TE-FMC) 111
Figure 3.4: Items from Which picture? activity (TE-FMC)... 113
Figure 3.5a: Feedback for Which picture? activity (TE-FMC).. 113
Figure 3.5b: Extra feedback for Which picture? activity (TE-FMC).. 114
Figure 3.6: Feedback for OVS items in Which picture? worksheet (TE-FMC).......................... 114
Figure 3.7: Items from Who's doing what? activity (TE-FMC) 115
Figure 3.8: Feedback for Who's doing what? activity (TE-FMC) 115
Figure 3.9: Items from Missing Words activity (TE-FMC) 116
Figure 3.10: Feedback for Missing nouns activity (TE-FMC) 117
Figure 3.11: Recap of grammatical rule (TE-F).. 118
Figure 3.12: Items from Picture Matching activity (TE-F) 119
Figure 3.13a: Feedback for Picture Matching activity (TE-F) 120
Figure 3.13b: Extra feedback for Picture Matching activity (TE-F) 120
Figure 3.14: Noticing task items from Picture Matching task (TE-F) 121
Figure 3.15: Feedback for computer-based Noticing task 122
Figure 3.16: Items from Sensible or silly? activity (TE-F) 123
Figure 3.17: Feedback from Sensible or silly? activity (TE-F).. 124
Figure 3.18: Items from Missing noun activity (TE-F) 124
Figure 3.19: Feedback for Missing noun activity (TE-F) 125
Figure 3.20: Items from Sentence Matching task (Set B) 131
Figure 3.21: Items from Gap-fill task (Set A)... 132
Figure 3.22: Items from Act-Out Comprehension task (Set B) 134
Figure 3.23: Items from Act-Out Production task (Set B) 135
Figure 3.25: Item from Sentence Reconstruction task (Set A).. 138
Figure 3.26: Item from Matching words task 143
Figure 4.1: Total scores on the vocabulary test 159
Figure 4.2: Total scores on the Sentence Matching task 160
Figure 4.3a: Scores on the SVO test condition (Sentence Matching) 162
Figure 4.3b: Scores on the OVS test condition (Sentence Matching) 163
Figure 4.4: Total scores on the Gap-fill task... 165

Figure 4.5a: Scores on the SVO+Subj test condition (Gap-fill).. 167
Figure 4.5b: Scores on the OVS+Subj test condition (Gap-fill) 168
Figure 4.5c: Scores on the SVO+Obj test condition (Gap-fill) 169
Figure 4.5d: Scores on the OVS+Obj test condition (Gap-fill).. 170
Figure 4.6: Total scores on the Act-Out Comprehension task 173
Figure 4.7a: Scores on the SVO test condition (Act-Out Comprehension 174
Figure 4.7b: Scores on the OVS test condition (Act-Out Comprehension).............................. 175
Figure 4.8: Total scores on the Act-Out Production task 177
Figure 4.9a: Scores on the Subject test condition (Act-Out Production)................................. 178
Figure 4.9b: Scores on the Object test condition (Act-Out Production).................................. 180
Figure 4.10: Total scores on the Sentence Repetition task 185
Figure 4.11a: Scores on the SVO+Subj test condition (Sentence Repetition) 186
Figure 4.11b: Scores on the OVS+Subj test condition (Sentence Repetition) 187
Figure 4.11c: Scores on the SVO+Obj test condition (Sentence Repetition) 188
Figure 4.11d: Scores on the OVS+Obj test condition (Sentence Repetition)........................... 189
Figure 4.12: Total scores on the Sentence Reconstruction task (Order) 201
Figure 4.13: Total scores on the Sentence Reconstruction task (Explanation) 202
Figure 5.1a: Distribution of scores on the written tasks (post-test).. 221
Figure 5.1b: Distribution of scores on the metalinguistic task (post-test) 221
Figure 5.1c: Distribution of scores on the oral tasks (post-test).. 222

Acknowledgements

I would like to extend my heartfelt thanks to my supervisor, Emma Marsden, for your encouragement and belief in me. Thank you also for the many fantastic opportunities you made available to me. I have learnt so much and am truly grateful for everything you have done.

I would like to thank my Thesis Advisory Panel supervisor, Danijela Trenkic, for your advice on the design of my study. Thank you also to Anna Capińska, David O'Reilly and Helen Saddler for spending several long days in school assisting me with data collection.

Thank you to the teachers and pupils who welcomed me so warmly into their schools. Without you, this study would not have been possible.

My thanks go to my fellow PhD colleagues for many stimulating discussions and in particular to Ali Bostancioglu for your friendship over the last four years, and for keeping me sane on a daily basis!

I would like to thank my family and friends for your unwavering support, and in particular my parents, Heather and Fionn Hanan, for teaching us that we can achieve anything we set our minds to. Thank you also to my parents-in-law (to be), Jane and Robert Kasprowicz, for making the first step in this journey possible.

Finally, to my fiancé, Richard, thank you for your endless love, patience and understanding, and for inspiring me, every day, to push myself to achieve my best. Thank you.

Author's declaration

All data presented in this thesis is original. All data collection and analysis was carried out by the author, Rowena Hanan, with the exception of the oral task data, which was collected with the assistance of Dr Emma Marsden, Anna Capińska, David O'Reilly and Helen Saddler.

Chapter 1: Introduction

1.1 The research context

Grammar teaching and learning has been a key area of inquiry within second language acquisition (SLA) research for many years. Numerous studies (for reviews, see R. Ellis, 1999; Norris & Ortega, 2000; Spada & Tomita, 2010; VanPatten, 2004c) have sought to determine the effect of explicit instruction on the acquisition of target language grammar; however there continues to be extensive debate as to whether instruction simply results in the learner developing explicit knowledge of grammatical rules (e.g. Hulstijn, 2002; Krashen, 1982; M. Paradis, 1994) or whether it can also impact the learners' implicit knowledge and underlying grammatical system (e.g. Anderson, 2005; DeKeyser, 2007; N. Ellis, 2005; Schmidt, 1990; VanPatten, 2004a).

Notably *child* SLA has received relatively little attention compared to that of adults. Despite the fact that observations based on children's first language acquisition are often the basis for theories regarding the processes involved in SLA (Philp, Mackey, & Oliver, 2008), child SLA has rarely been studied as a "subfield" in its own right (J. Paradis, 2007, p. 387). Correspondingly relatively few studies have investigated the efficacy of explicit grammar instruction for young learners, arguably due to the fact that they are thought to have access to an implicit language learning mechanism (Lenneberg, 1967). It is important to note, however, that within the foreign language classroom context, learners may not be able to capitalise on their ability to learn language implicitly, since exposure to the target language is substantially limited (Gass & Selinker, 2008; Muñoz, 2006, 2008b). Therefore it is important to explore whether young learners may benefit from more explicit instruction when learning within such an environment.

In situations where input alone is not sufficient, a type of focus-on-form which aims to improve learners' processing of L2 forms through exposure to more structured input may be optimal (Wong, 2004a). Indeed, Doughty (2003) proposed that "the goal of L2 instruction should be to organise the processing space to enable [learners] to notice the cues *located in the input*" (p. 298). Furthermore, for children the teaching of abstract rules may not be effective, rather a more appropriate approach might be to employ activities in which the task demands themselves require the learner to attend to the relevant second language feature (Harley, 1998). As such the present study will compare two types of input-based instruction; TE-FMC and TE-F. Both interventions will provide the learners with explicit information relating to the target feature followed by listening and reading activities in which attention to either the target grammatical form *and* its function within the input (TE-FMC), or the target grammatical form only (TE-F) are made task-essential.

The present study, therefore, seeks to contribute to effect of instruction research by investigating the extent to which explicit teaching of foreign language grammar is effective for young learners learning within the foreign language classroom environment. To this end the study addresses the following research questions (RQs):

1) Does explicit grammar instruction improve young learners' a) comprehension and b) production of the target grammatical feature?

2) To what extent does explicit grammar instruction develop different types of knowledge of the target grammatical feature?

3) Following explicit information, is intentional practice in attending to the target form-meaning connection more beneficial than intentional practice in attending to the target grammatical form only?

1.2 The educational context: Overview of the state education system in England

Compulsory, full-time education runs from ages 5 to 16 in England[1]; pupils attend primary school from ages 5 to 11, and secondary school from ages 11 to 16. Primary education can be subdivided into two key stages: Key Stage 1 (KS1) for children aged 5 to 7 (Years 1 and 2); and Key Stage 2 (KS2) for children aged 7 to 11 (Years 3 to 6). Secondary education consists of Key Stage 3 (KS3) for pupils aged 11 to 14 (Year 7 to 9) and Key Stage 4 (KS4) for pupils aged 14 to 16 (Years 10 and 11). The national curriculum is aligned with each of the key stages and sets out the subjects to be taught at each level (see DfE, 2013b). At primary level, the KS1 and KS2 curriculums comprise three core subjects (English, Mathematics, Science) and seven foundation subjects (Art and Design, Computing, Design and Technology, Geography, History, Music, Physical Education). From September 2014 foreign languages were introduced as an additional compulsory foundation subject at KS2 (DfE, 2013c). Pupils sit Standard Assessment Tests (SATs)[2] in numeracy, English literacy (i.e. reading), and English grammar, punctuation and spelling at the end of KS2 (age 11). Pupils' abilities in Mathematics, English and Science are

[1] https://www.gov.uk/know-when-you-can-leave-school (Accessed 3rd February 2015)
[2] https://www.gov.uk/government/publications/key-stage-2-assessment-and-reporting-arrangements-ara/the-national-curriculum-tests (Accessed 4th February 2015)

usually assessed 'in-house' at the end of KS1 in relation to the national curriculum level descriptions[3].

1.3 Changes to foreign language policy in England

The last few decades have witnessed extensive changes to foreign language provision within the primary (and secondary) education sectors in England. Crucially, in the National Curriculum 2000 (DfEE & QCA, 1999) foreign languages were included for the first time as an optional, non-statutory subject with an accompanying scheme of work. In 2002 the government launched the National Languages Strategy "Languages for all: Languages for life", which proposed that *all* children should be given the opportunity to study a foreign language whilst at primary school. Notably, this renewed interest in foreign language teaching at primary schools was fuelled in part by the pervading perception that in terms of language learning earlier equals better (The Nuffield Languages Inquiry, 2000). Indeed, the then Prime Minister Tony Blair claimed that "Everyone knows that, with languages, the earlier you start, the better"[4]. In relation to the foreign language classroom context, however, the research evidence is not as clear cut as this would suggest (see section 2.1).

In response to these policy changes, numerous initiatives were undertaken to explore the nature of foreign language provision at primary school (e.g. Driscoll, Jones, & Macrory, 2004; Marilyn Hunt, Barnes, Powell, Lindsay, & Muijs, 2005; Muijs et al., 2005; The Nuffield Languages Inquiry, 2000; Wade, Marshall, & O'Donnell, 2009). Throughout the 2000s a steady increase was observed in the percentage of primary schools offering a foreign language at KS2 (to 92% in 2008) (Wade et al., 2009). Notably, these surveys also highlighted a number of recurring challenges to the potential for maintaining and/or increasing language provision, such as limited time, other curriculum priorities, lack of staff expertise and confidence.

In the late 2000s, reviews of the National Languages Strategy (Dearing & King, 2007) and the primary curriculum (Rose, 2009) concluded that foreign languages should be made a statutory requirement for KS2. Subsequently the call for evidence of the 2011 National Curriculum Review set forth considerable support for the inclusion of foreign languages as a compulsory KS2 foundation subject, with 82% of respondents (*N* = 2276) in favour of such a change (DfE, 2011, p. 41). In response to these calls, foreign languages were included in the National Curriculum (DfE, 2013c) for the first time as one of eight compulsory foundation subjects. The 2013/2014 Language Trends survey (Board & Tinsley, 2014) revealed that, in preparation for the introduction of the new curriculum, 95% of primary schools were teaching

[3] https://www.gov.uk/government/collections/statutory-guidance-schools#curriculum-until-july-2015 (Accessed 4th February 2015)

[4] Tony Blair, Romanes Lecture, Oxford, December 1999, quoted in (The Nuffield Languages Inquiry, 2000)

a foreign language, with a majority of schools providing 30 to 45 minutes per week. Since September 2014 all primary schools within England have been required to provide instruction in a foreign language of their choice at KS2.

1.4 Foreign language teaching at KS2

The current KS2 curriculum for foreign language teaching provides relatively brief guidelines on the teaching of this newly compulsory subject. First, the curriculum states that teaching should provide a balance of both spoken and written input in order to develop the four key skills: listening, speaking, reading, writing (DfE, 2013d, p. 1). Secondly the guidelines emphasise that learners should engage in conversational use of the target language, as well as comprehend and present ideas and information in both oral and written format. The use of stories, songs, poems, and rhymes is encouraged as a means of improving learners' vocabulary, pronunciation, and understanding of the patterns and sounds of the language. The curriculum therefore promotes a "competency-based" approach to language learning. Nevertheless, it is important to note that no 'top-down' indication of the linguistic aims (i.e. vocabulary, grammar) to be achieved by the end of KS2 are included.

Despite the emplahsis on both oracy and literacy within the new foreign language curriculum, in practice they are not always given equal weighting (Mitchell, 2011). Indeed following their inspection of ten Pathfinder[5] local authorities, OFSTED (2005) observed that pupils tended to have developed confidence in speaking in the target language and good listening skills; however literacy (i.e. reading, writing) skills were underdeveloped, and few pupils had an understanding of how different languages work. Further in their exploration of the nature of foreign language provision within 40 primary schools in England, Cable et al. (2012) found that a majority of teachers were employing "fun" oracy-based activities (e.g. songs, game-like activities, role plays) designed to promote listening and speaking abilities, but markedly less attention was paid to developing target language literacy. Within the most recent Language Trends survey, many teachers claimed that "currently we focus on helping our children become confident at speaking and listening…" (p. 49). Notably, this skew towards a primarily oracy-based approach may reflect the fact that children's L1 literacy is still developing (Board & Tinsley, 2014; Cable et al., 2012; Marilyn Hunt et al., 2005), as well as factors such as lack of staff expertise and confidence in teaching a new sound-spelling system, and the limited time dedicated to language teaching.

Given the emphasis on both oracy and literacy within the curriculum guidelines, Graham, Courtney, Marinis, and Tonkyn (2014) carried out a study to investigate the

---

[5] From 2003 to 2005 19 Pathfinder local authorities were set up in order to trial a range of language teaching approaches across approximately 1,000 schools (Muijs et al., 2005)

differential benefits (linguistic and motivational) of these two approaches for primary pupils' learning of French during the transition phase from primary to secondary school. Relatively little research has sought to assess the impact of these teaching approaches on young learners' attainment, therefore "little is known about the kind of learning in the primary school that best prepares learners for further language study in the context of England" (Graham et al., 2014, p. 3). Graham et al. (2014) observed a steady, significant, albeit small, amount of improvement in the learners' performance across a range of language tasks[6] over the two year study. No differences in attainment were observed between the two teaching approaches, although the literacy-based approach was found to be more beneficial for those learners with low L1 literacy. Based on their findings, Graham et al. (2014) proposed that carefully designed literacy-based activities should be incorporated within language teaching alongside those tasks which focus on the development of oral communication abilities. Nevertheless it is important to note that no control group was included, therefore it is not possible to eliminate the potentially confounding effect of history, maturation, test effect (see section 3.1.2.2) on the learners' performance over the two year study. Further research is therefore needed to explore in more depth appropriate teaching approaches for this context.

1.5 Foreign language grammar teaching at KS2

1.5.1 Curriculum guidelines and current practice

Alongside the more competency-based foci of the current curriculum, and of primary interest for the present study, knowledge of target language grammar is also promoted to a certain extent. Pupils should be taught to:

> "understand basic grammar appropriate to the language being studied, including (where relevant): feminine, masculine and neuter forms and the conjugation of high-frequency verbs; key features and patterns of the language; how to apply these, for instance, to build sentences; and how these differ from or are similar to English" (DfE, 2013d, p. 3).

Nevertheless, within the Language Trends survey, "the teaching of reading including the understanding of grammar" (73% respondents), and "the teaching of writing including the correct use of grammar" (77% respondents) were identified as two of the biggest challenges and areas where a majority of schools felt least strong (Board & Tinsley, 2014, p. 44).

In terms of how grammar is taught, given the recency of the introduction of foreign languages into the primary curriculum, no 'standard' textbooks as yet exist for use in schools nationwide (contrary to the common practice at KS3 and 4). Rather many schools have tended to rely on government published schemes of work (QCA, 2007a, 2007b, 2007c), as well as

[6] Language tasks focussed on vocabulary and grammatical knowledge, e.g. gender, article-noun agreement, noun-adjective agreement, subject-verb agreement (Graham et al., 2014, p. 7)

commercially-available (e.g. Cheater, 2007; Neubauer, Pearson, & Whittle, 2015; Seccombe, 2014a, 2014c) and freely-available resources created by teachers for teachers (Minette, Crellin, & Holden, 2009; Seccombe, 2014b; see also TES website)[7]. In accordance with the curriculum guidelines such schemes of work incorporate the teaching of key grammatical concepts (e.g. gender, agreement, case marking, pronouns, negation, question formation, word order etc.). Further an examination of a selection of these KS2 resources, as well as a selection of textbooks currently utilised at KS3 and KS4 (e.g. Edexcel GCSE German, Lanzer & Wardle, 2009; Zoom Deutsch 1, Schicker, Waltl, & Malz, 2011), revealed that grammar teaching often tends to utilise the following activities, or a combination thereof:

- listening / reading activities in which the input is 'enriched' with examples of the target feature but in which the focus of the activity is on meaning (e.g. vocabulary learning / practice)
- identification of the target feature within enriched input activities
- presentation / discussion of the rules governing use of the target feature
- practice (spoken or written) in producing the target feature in discrete, closed-item tasks

Due to the lack of standardized teaching materials for KS2, there is "a wide spectrum of practice and lack of consistency between schools in approach and outcomes achieved" (Board & Tinsley, 2014, p. 10). It is also important to consider that the effectiveness of such activities for grammar learning, and for learners at different ages, is unclear (see Chapter 2). Indeed in their investigation of language teaching for learners starting at 5, 7, or 11 years old, Myles and Mitchell (2012) observed that it was the older starters who made greater gains in terms of grammatical knowledge following instruction based on the current framework for languages. As proposed by Graham et al. (2014), teaching approaches and materials *should* be designed on the basis of research evidence, yet very often are not. Therefore, classroom-based research is needed in order to investigate *how* grammar can be most effectively taught within the primary context and for learners at different stages of primary education.

1.5.2 Promoting language awareness

In line with the current curriculum, within the KS2 framework for languages (DfES, 2005) 'knowledge about language' and 'language learning strategies' are identified as two additional learning tools, which are important in and of themselves. Knowledge about language is defined as understanding of how languages work and awareness of the rules and patterns within a

[7] TES website contains free resources, created and uploaded by teachers for all curriculum subjects at KS1-5, see https://www.tes.co.uk/primary-teaching-resources/

language(s) (DfES, 2005). Further, the framework proposes that learners can draw on their developing explicit understanding of the target language, as well as prior knowledge of their L1, in order to create new language and adapt their language use for different contexts (i.e. formal or informal). In addition the framework promotes the development of learners' familiarity with language learning strategies, which can be utilised in the learning of any foreign language. These cross-cutting strands (knowledge about language, language learning strategies) also have implications for the issue of transition and the discrepancy which often occurs between the language(s) learnt at primary school and the language(s) taught at secondary school (Board & Tinsley, 2014; Cable et al., 2012; Marilyn Hunt et al., 2005; Mitchell, 2011; Wade et al., 2009). As a consequence the importance of improving young learners' knowledge and understanding of how languages work and the differences between languages, i.e. their language awareness had been emphasised (Cable et al., 2012; Hawkins, 2005; Marilyn Hunt et al., 2005; Mitchell, 2011).

Language awareness can be defined as "explicit knowledge about language, and conscious perception and sensitivity in language learning, language teaching and language use" (ALA, 2012). For many years, Hawkins (e.g. 1999, 2005) promoted the use of language awareness programmes as part of a 'language apprenticeship' to provide learners with the skills necessary to learn a foreign language, push learners to ask questions about language and look at language objectively. A number of other initiatives have been evaluated in recent years such as *Discovering Language*[8] (Barton, Bragg, & Serratrice, 2009), which offers a multilingual language awareness programme, and *Springboard to Languages*[9] (Roehr, 2012; Tellier, 2012b), which develops language awareness through the teaching of Esperanto[10]. The evaluations of such programmes have revealed that they are successful in increasing learners' awareness of grammatical structures and understanding of grammatical meta-language, as well as the differences between English (the L1) and foreign language structures (Barton et al., 2009; Roehr, 2012). Similarly Tellier and Roehr-Brackin (2013b) investigated the metalinguistic awareness of pupils who had received instruction in Esperanto plus another European language (typically French or Spanish) at primary school, compared to those who had received instruction in European languages only. Whilst observing no overall differences between the two groups' on a series of metalinguistic tasks, the performance of the Esperanto group was found to be more homogenous. Tellier and Roehr-Brackin (2013b) proposed that the teaching of a low difficulty language such as Esperanto may be helpful in developing both low and high ability learners' metalinguistic awareness, and in fostering learners' capacity for explicit

[8] http://sha.org.uk/Home/About_us/Projects/Discovering_language/Discovering_language/
[9] http://www.springboard2languages.org/
[10] Esperanto is a language made up of 917 root words and 16 key rules of grammar, constructed by Dr L. L. Zamenhof. The grammar is simple, transparent and free from irregularities (Tellier, 2012a).

language learning, which tends to be characteristic of the learning which takes place within the foreign language classroom (see section 2.1).

An additional theme which arises in the discussion of the role of language awareness is its use as a tool for developing learners' knowledge of their L1 (in this context English), as well as a means of bridging the gap between the L1 and L2 teaching within the curriculum (Barton et al., 2009; Hawkins, 1999, 2005; Martin, 2000). Martin (2000) argues that foreign language work should be "explicitly associated with English language work" (p. 7) and that one can reinforce the other. Young learners' developing L1 literacy, for example, could serve to aid learners' understanding of the relationship between sounds and writing in both the L1 and L2, and vice versa. Further Cable et al. (2012) observed that even for those pupils who encountered difficulty with English literacy, foreign language lessons helped pupils feel more assured and increased confidence. Similarly Graham et al. (2014) found that a literacy-based approach to foreign language teaching was particularly beneficial for those learners whose L1 literacy was low. As such it has been proposed that foreign language teaching should be incorporated into a broader cross-curricular programme, explicitly-linked to the L1 (as well as other languages), rather than viewed as a "bolt-on experience" (Cable et al., 2012; Martin, 2000). Indeed the KS2 framework states that "when learning a new language, children reinforce and reinterpret knowledge and understanding gained in learning their first language(s)" (DfES, 2005, p. 9). Further the national curriculum for English (DfE, 2013a) emphasises the importance of children developing awareness of language and L1 grammar from as early as age 5. Similarly in the English grammar, punctuation and spelling component of the KS2 SATs test, pupils are required to demonstrate their knowledge and understanding of key grammatical concepts. Developing an awareness and understanding of foreign language grammar could reinforce this knowledge of the learners' L1. This is of particular relevance for the present study, which deals with concepts (subject, object) included in the primary English curriculum.

The national curriculum, then, advocates a fairly explicit approach to language learning (L1 and L2), which includes the provision of instruction in target language grammar. Such an approach is arguably inconsistent with the view that young learners are able to learn languages implicitly and therefore benefit most from extensive exposure to the target language (e.g. DeKeyser & Larson-Hall, 2005; Lenneberg, 1967; Muñoz, 2008b), rather than explicit attention to form (see section 2.1). It is also important to note, as acknowledged by Cable (2012) that the components of the national curriculum and associated frameworks "have been devised a priori and are not underpinned in detail by empirical evidence on progression and learning outcomes" (p. 367). Therefore classroom-based research within the context of England is necessary in order to determine the benefits of the teaching approaches advocated within the current guidelines. The present study will seek to address one issue of relevance, namely the efficacy of grammar instruction for young learners within the foreign language classroom context in England. Further, in line with the curriculum guidelines for KS2 languages, the present study:

- incorporated outcome measures related to the four key skills (reading, writing, listening, speaking)
- incorporated both written and aural stimuli in intervention activities
- promoted learners' awareness of the target grammatical feature and related concepts (in the L1 and L2)
- trained learners in utilising a language learning strategy (attention to the target grammatical feature) to aid comprehension of target language input

In addition, given the substantial variation in language teaching provision that often occurs between individual schools (Board & Tinsley, 2014; Cable et al., 2012), all classes received pre-teaching in the target language from the researcher prior to participating in the study (see section 3.3.3).

1.6 Outline of the thesis

Having established the educational context in which the present study is set in Chapter 1, Chapter 2 will present a review and critique of the relevant research literature, divided into three sections: the role of age in SLA and the effectiveness of explicit grammar instruction for child L1 and L2 learning; the theoretical standpoints on the nature and role of explicit knowledge in language acquisition; the rationale for and efficacy of input-based explicit grammar instruction for SLA. Chapter 3 will present the methodology and methods utilised in the study. Chapter 4 details the results for each of the outcome measures, both over Time (pre-, post-, delayed post-test) and Between-group (TE-FMC, TE-F, Control). Chapter 5 presents the analysis exploring the relationship between the learners' performance across the outcome measures, a more fine-grained analysis of individual learners' performance, and analysis exploring any confounding effects of grammatical sensitivity. Chapter 6 presents a critical discussion of the findings of the study in relation to the three research questions that the study sought to address. Finally Chapter 7 summaries the study and main findings and reviews the limitations of the study. In conclusion, the implications for the primary foreign language classroom and the contributions of the study to language learning research are considered.

Chapter 2: Literature Review

2.1 How do children acquire a second language?

2.1.1 Age effects in SLA

The proposal that there is a critical age for language learning has received considerable attention within second language acquisition (SLA) research. Consequently a multitude of studies have investigated differences between child and adult language acquisition. Whilst a difference has consistently been observed in terms of rate or level of ultimate attainment (e.g. DeKeyser, 2000; Johnson & Newport, 1989; Muñoz, 2008b; Snow & Hoefnagel-Höhle, 1978), there continues to be extensive debate as to the reasons for the observed differences (Philp et al., 2008). On the one hand proponents of the Critical Period Hypothesis (Lenneberg, 1967) have argued that there is a qualitative difference between the ways in which children versus adults acquire a second language (e.g. DeKeyser, 2000; DeKeyser & Larson-Hall, 2005; Hyltenstam & Abrahamsson, 2003). On the other hand alternative explanations have been put forward (e.g. Competition Model, MacWhinney, 2001), which suggest that observed differences are due to the level of exposure the learner has had to another language, rather than there being any inherent difference between the acquisition processes involved.

Whilst the present study is not primarily concerned with comparing child and adult language acquisition, it is nevertheless important to explore the various theories put forward as to how children are thought to acquire a second language.

2.1.2 Critical Period Hypothesis

The proposed distinction between how children and adults acquire language was first characterised by Penfield and Roberts (1959) and subsequently by Lenneberg (1967) as a maturational change occurring in the brain before puberty (age 12 to 13). Lenneberg (1967) refers to a *critical period* for language acquisition during which children have an innate capacity for language learning and the ability to learn languages implicitly. However, "automatic acquisition from mere exposure seems to disappear after this age", i.e. post-puberty (Lenneberg, 1967, p. 176). Based on these claims the Critical Period Hypothesis (CPH) was proposed, which states that acquisition from mere exposure, sometimes referred to as implicit learning, is the only language acquisition device available to the young learner, but is severely limited in older learners and adults (DeKeyser & Larson-Hall, 2005). Older learners and adults are therefore reliant on explicit language learning mechanisms (DeKeyser, 2003). Consequently language learning post-critical-period has been described as requiring conscious and laboured effort (Lenneberg, 1967, p. 176). It has also been suggested that there may be multiple 'sensitive periods', rather than one critical period, since the changes that are thought to occur in language learning ability may vary in onset and effect (Long, 1990). For example the sensitive period for phonological development may close as early as age six, whereas for morphology and syntax, the sensitive period may last up until age 15 (Long, 1990).

The CPH therefore has consequences for both L1 acquisition and child and adult L2 acquisition. In terms of L2 acquisition, the CPH has implications for the level of attainment achievable at different starting ages, positing that only those learners who begin learning during the critical period are able to reach native-like (or ultimate) attainment in the L2 (DeKeyser & Larson-Hall, 2005; Hyltenstam & Abrahamsson, 2003). Notably it is important to make clear the distinction between ultimate attainment and rate of acquisition (Krashen, Long, & Scarcella, 1979). Whilst younger learners are thought to have an advantage in terms of ultimate attainment, older learners have been shown to acquire language at a faster rate in the initial stages, since their superior cognitive development enables the use of faster explicit learning mechanisms (e.g. Muñoz, 2008b; Snow & Hoefnagel-Höhle, 1978). Consequently the CPH may be of less consequence for rate of acquisition (DeKeyser & Larson-Hall, 2005; Hyltenstam & Abrahamsson, 2003; Long, 1990).

Support for the CPH has been underpinned by research findings which have provided evidence that learners, who begin learning a second language in adulthood, are unable to reach ultimate attainment in that language even after an extensive period of time (DeKeyser & Larson-Hall, 2005). Numerous studies have been carried out in order to investigate the relationship between age of acquisition (AoA) and L2 proficiency in both morphosyntax and phonology (pronunciation) (for reviews, see Birdsong, 2005; DeKeyser & Larson-Hall, 2005; Hyltenstam & Abrahamsson, 2003; Muñoz & Singleton, 2011). One of the most influential studies was that of Johnson and Newport (1989) in which Chinese and Korean immigrants in the USA were tested on their knowledge of basic grammatical structures in English via a grammaticality judgement task (GJT). Johnson and Newport's findings revealed a significant negative correlation between AoA and L2 proficiency until puberty (age 16). Post-puberty, a much weaker relationship was observed and performance was found to be low and highly variable. Similar results have been observed in replications of Johnson and Newport's seminal study, with a decline in proficiency up to a certain age, which levels-off for older learners and adults (e.g. DeKeyser, 2000; Johnson, 1992). Such findings have been taken to support the existence of a critical period for language acquisition and as evidence that children are 'better' language learners, with respect to eventual outcome, due to their ability to learn language implicitly.

Nevertheless there remains extensive debate as to the reasons for the observed disparity between child versus adult SLA and whether maturational constraints are accountable (DeKeyser & Larson-Hall, 2005; Hyltenstam & Abrahamsson, 2003; Muñoz & Singleton, 2011). Criticisms have arisen in relation to methodological issues; for example the participants in Johnson and Newport's (1989) study had been residing in the USA for a minimum of five years, which it is argued is unlikely to be a sufficient amount of time for the learners to have reached ultimate attainment (Bialystok & Hakuta, 1994, cited in Hyltenstam & Abrahamsson, 2003). Additionally studies have presented evidence of adult acquirers achieving native-like proficiency in the target language (e.g. Birdsong & Molis, 2001; L. White & Genesee, 1996). Such findings have been taken to falsify the CPH and the existence of maturational constraints on language acquisition (Hyltenstam & Abrahamsson, 2003). Further it has been argued that rather than maturational constraints, factors such as amount and quality of input, opportunities to practice, learners' attitudes, and the context can account for the variation in attainment between child and adult learners (DeKeyser & Larson-Hall, 2005; Muñoz & Singleton, 2011).

2.1.2.1 Theories of child language acquisition

The CPH was put forward based on the observed link between maturational changes found to occur in the brain around puberty and key milestones in language acquisition (Lenneberg, 1967). However it has been proposed that general maturational changes may not sufficiently explain the observed age effects in language acquisition. Rather there may be deeper causes, for example, a qualitative change in the language learning capacities available to younger versus older learners (DeKeyser & Larson-Hall, 2005, p. 99).

One theory put forward, within the framework of Universal Grammar (UG), is Bley-Vroman's Fundamental Difference Hypothesis (FDH). In the original formulation of the hypothesis Bley-Vroman (1990) argues that there is a fundamental difference between child and adult language acquisition; child language acquisition is guided by an innate "domain-specific language acquisition system" (p. 13). However this system ceases to be operative in adults, who are only able to access more general learning procedures. The FDH posits that a child's innate language acquisition system consists of two parts; Universal Grammar, "a definition of possible grammar", and a set of Learning Procedures, "a way of arriving at a grammar based on available data" (Bley-Vroman, 1990, p. 14). Bley-Vroman (1990) refers to a number of age-related characteristics of language acquisition as support for these claims, for example the rarity of cases of "complete success" in adult SLA, the negative correlation between age and attainment, the importance of instruction for older learners, and the role of affective factors in shaping adult SLA. DeKeyser (2000, 2003) characterised this distinction in terms of implicit and explicit learning; namely that the implicit learning mechanisms which drive child language acquisition are no longer available for adult learners, who must therefore utilise explicit, problem-solving mechanisms.

In response to advances in UG and language learning research, a reformulation of the FDH has since been put forth, in which Bley-Vroman (2009) proposes that, rather than a distinction between "domain-specific" and "domain-general" processing, the difference between child and adult language acquisition can be defined in terms of a reliance on "grammar driven processing" (child) versus "shallow structure processing" (adult). In line with the proposed existence of two language processing mechanisms, the Competition Hypothesis, an alternative UG-based theory put forward by Felix (1985), claims that child language acquisition is guided by a "language-specific module", whereas in adult SLA this module is in competition with the more general "problem-solving module". Similarly Krashen (1982) argues that the ability to acquire language implicitly does not disappear post-puberty, rather the same natural language acquisition device is active in adults as well as children (p. 10).

Alternative usage-based theories have also been proposed. Tomasello (2000) highlights a shortcoming of UG-based accounts in that the assumption is often made that children possess a fully developed adult-like representation of grammar. Contrary to this assumption Tomasello (2000) argues that child language acquisition is item-based and piecemeal, rather than being guided by "system-wide syntactic categories" and parameters. In other words, rather than being driven by underlying linguistic competencies, children acquire language based on the input they are exposed to. Consequently it is argued that child language acquisition is underpinned by more general cognitive abilities (Tomasello, 2000). Alternatively the Competition Model (MacWhinney, 2005) proposes that fundamentally L1 and L2 acquisition are reliant on similar processes, however differences arise due to the influence of a learner's experience with another language.

An important consideration relevant to such theories is the fact that, commonly, such theories tend to be based on a comparison of *child first* language and *adult second* language acquisition. Notably, however, the acquisition of a *second* language during childhood is likely to be distinct from both child L1 and adult L2 learning (Nicholas & Lightbown, 2008). Firstly, for the child acquiring a second language, the first language (L1) has already been established, nevertheless it has not become as entrenched as for adult learners (MacWhinney, 2005; Philp et al., 2008). Consequently the L1 is likely to have a differential effect depending upon the point during childhood at which the second language is acquired (MacWhinney, 2005). Additionally more general cognitive, problem-solving abilities may also have a role in child SLA; as children develop cognitive maturity, these abilities may play an increasingly important role (DeKeyser & Larson-Hall, 2005). Indeed Newport (1990) proposes that "language learning declines over maturation precisely because cognitive abilities increase" (p. 22). Finally it is important to consider that child SLA is "shaped by dynamic interactions of multiple factors" (Jia & Aaronson, 2003, quoted in Philp et al., 2008, p. 10), including cognitive, social, and cultural variables, the interaction of which make child SLA distinct from both child L1 and adult L2 acquisition. This section has set out a number of theories regarding the distinction between language acquisition in child- and adulthood. Building on what is known about child language acquisition, the next section will consider the assumption that in terms of language learning, younger is better.

2.1.2.2 Does younger equal better?

The CPH and observed age effects in language acquisition have fuelled claims that in terms of language learning, younger equals better. As detailed in section 2.1.2, previous research (e.g. DeKeyser, 2000; Johnson, 1992; Johnson & Newport, 1989) has demonstrated that younger learners appear to have a long-term advantage in terms of ultimate attainment, although opinions differ as to the role of maturational causes for this advantage. In contrast older learners and adults are found to be at an advantage in terms of the initial rate of learning they are capable of achieving (Krashen et al., 1979; Muñoz, 2008a).

However an important, yet often overlooked, issue in research investigating age-effects in SLA is the setting in which language learning occurs (Muñoz, 2006, 2008b). A majority of the research in this area has tended to be carried out in naturalistic, or immersion, settings, in which the language learner is immersed in the second language environment. Within this context the younger learner has been found to have a long term advantage (Muñoz, 2008b). Of issue is the fact that this finding from immersion settings (i.e. for ultimate attainment younger does equal better) has tended to be overgeneralised to the instructed setting (Muñoz, 2006). However within the foreign language classroom it is the older learners who are found to progress further and faster (Cenoz, 2003; DeKeyser, 2003; Muñoz, 2006, 2008b). Indeed Krashen et al (1979) observed that in the case of foreign language learning within the instructed setting, it is those learners who begin instruction later, who overtake those learners who begin at a younger age.

Although there is extensive debate as to the nature of the processes underlying child (and adult) acquisition, it is generally agreed that young learners are able to learn language implicitly (i.e. without conscious attention and awareness) (DeKeyser, 2003; DeKeyser & Larson-Hall, 2005). It is this ability to learn implicitly, which it is argued gives younger learners a long-term advantage when learning within the immersion setting (DeKeyser, 2003; Muñoz, 2008b). However, implicit language learning requires a huge amount of exposure to the target language, yet within the instructed setting exposure to the target language is substantially limited (Gass & Selinker, 2008). Particularly within the UK context, young learners will receive on average one hour of instruction in the target language (Cable et al., 2012; M. Hunt, Barnes, Powell, Lindsay, & Mujis, 2005; Wade et al., 2009), and there is a high incidence of L1 use (Muñoz, 2008b). Consequently within the instructed setting younger learners are deprived of their 'advantage' due to a lack of exposure to the target language (Muñoz, 2006). Exposing young learners to sufficient amounts of target language input in order to enable implicit acquisition processes to be employed is only possible within an immersion classroom, and not within a setting offering at most one hour of foreign language teaching per week (DeKeyser, 2000). As argued by DeKeyser and Larson-Hall (2005) "the observation that earlier is better only applies to certain kinds of learning, which schools typically cannot provide" (p. 88). The setting in which learning takes place and the quality of target language input to which the learner is exposed can therefore make a crucial difference to the cognitive processes available to children in language learning.

It has been argued that for adults engaging in language learning their reliance on implicit learning mechanisms is substantially reduced, since they have access to faster explicit, analytical processes (Bley-Vroman, 2009; DeKeyser, 2000). Utilising their ability to learn explicitly, enables adult language learners to make shortcuts in the learning of key grammatical structures (DeKeyser & Larson-Hall, 2005). Similarly the question arises, if children are not able to utilize their optimal (implicit) language learning system within the primary school context, can explicit knowledge and instruction potentially aid young learners' language acquisition within this environment? Consequently this study aims to investigate the effectiveness of explicit instruction for younger learners, learning within the instructed setting.

2.1.3 Child language acquisition and explicit grammar instruction

The role of explicit knowledge and the effectiveness of explicit instruction for language learning have primarily been investigated with regard to adult rather than child SLA (Bouffard & Sarkar, 2008). Nevertheless a number of studies have demonstrated the beneficial role of explicit instruction in aiding both child L1 and L2 acquisition.

2.1.3.1 …in the L1

Prior to addressing the question of whether explicit knowledge and grammar instruction may be useful for young learners learning a *foreign* language, it is important to first consider whether learners are able to make use of explicit grammatical knowledge in their L1. Primary-school education within the UK will commonly include explicit instruction relating to the grammar of the pupils' L1 (J. White, 2008). Indeed the current curriculum for KS2 English states that all pupils should acquire "an understanding of grammar and knowledge of linguistic conventions for reading, writing and spoken language" (DfE, 2013a, p. 3).

Predominantly, research has demonstrated that linguistic (i.e. morpho-syntactic) awareness can positively affect aspects of L1 reading and writing, in particular spelling accuracy (e.g. Bryant, Nunes, & Bindman, 2000; Carlisle, 2000; Deacon & Bryant, 2005; Nunes, Bryant, & Olsson, 2003). It has been argued that successfully learning to read and spell depends to a considerable extent on the child's knowledge and explicit awareness of grammar (Bryant, Devine, Ledward, & Nunes, 1997).

A longitudinal research study by Bryant et al (2000) investigated the relationship between children's (eight to ten years old) explicit linguistic awareness and their understanding of the orthographic rule governing the use of the apostrophe used to denote possession. The participants completed a series of awareness tasks (morpho-syntactic, phonological, syntactic/semantic) followed by a spelling task in which they had to write singular nouns in the genitive case (with apostrophe) or plural nouns in the nominative/accusative cases (without apostrophe). Bryant et al (2000) found that success in learning correct use of the apostrophe depended upon the learners' explicit morpho-syntactic awareness.

Similarly, Bryant et al (1997) sought to determine whether instruction would be effective in improving learners' knowledge of the grammatical function of apostrophes. Two intervention studies were conducted with children aged nine to 11. The experimental group in both studies received instruction relating to the use of apostrophes with genitive nouns and their performance was compared to that of a taught control group (same materials but no explicit instruction relating to the use of apostrophes), and an untaught control group. In both studies the experimental group was found to significantly improve in their use of the target feature compared to either control group (Bryant et al., 1997). Further the second study found additional evidence that it was the children's explicit awareness of grammatical distinctions which played an important role in learning about apostrophes and that the difficulties the learners had in using apostrophes were due to limited awareness of the genitive case (Bryant et al., 1997). Further, in line with such findings, a study by Nunes et al (2003) demonstrated that instruction resulted in significant gains in participants' (aged seven to eight) use of morphological spelling rules (e.g. how morpheme boundaries affect the pronunciation of particular letter sequences, such as *sh* in *misheard* or *disheartened*), as measured on a standardised read-aloud test and a spelling assessment.

In addition to studies investigating the role of morpho-syntactic awareness, two in-depth systematic reviews were carried out into the effectiveness of teaching a) syntax (sentence-level grammar) (Andrews, Torgerson, Beverton, Locke, et al., 2004) and b) sentence combining (e.g. use of conjunctions) (Andrews, Torgerson, Beverton, Freeman, et al., 2004) on learners' accuracy in written comprehension. In a review of 18 studies, Andrews, Torgerson, Beverton, Freeman et al. (2004) found that grammar teaching methods such as *sentence combining*[11] were effective in improving the syntactic maturity of learners from as young as age 5 to age 16. In contrast Andrews, Torgerson, Beverton, Locke et al (2004) reviewed 10 studies, which had explored the effectiveness of traditional (e.g. classifying and describing the relationship between internal elements of a sentence) and / or transformative (e.g. teaching the basic deep structural rules and how they transform into actual spoken or written utterances) grammar teaching approaches and argued based on their review that no convincing evidence has yet been put forward to suggest that the teaching of syntax is useful in improving learners' accuracy in writing.

Overall, an examination of the current research into the role of explicit knowledge and morpho-syntactic awareness in L1 learning suggests that "explicit instruction about morphemes is helpful to children's learning" (Hurry et al., 2005, p. 187), although teaching focussed on sentence-level syntax may not be effective (Andrews, Torgerson, Beverton, Locke, et al., 2004). Further the studies presented above have demonstrated that young learners do have a certain amount of explicit knowledge about language in their L1 and that a learner's level of morpho-syntactic awareness can have an impact on their ability to understand the function of, and correctly utilise, certain linguistic features, such as the apostrophe to mark possession (Bryant et al., 2000).

One final consideration worthy of note is that research investigating learners' linguistic awareness in their L1 is primarily concerned with the impact of this explicit knowledge on learners' use of key features in 'offline' tasks such as reading and writing. When completing such tasks learners have time to access and utilise the relevant explicit knowledge. In SLA research, however, it is also important to consider whether linguistic awareness can be instrumental in aiding the acquisition of key linguistic features for use in 'online' tasks, such as spontaneous oral communication. With this in mind, the next section will outline research which has been carried out into the role of explicit knowledge and instruction in SLA.

2.1.3.2 ...in the L2

The potential benefits of employing consciousness-raising tasks with children learning a second or foreign language might seem minimal, given the general assumption that children learn language implicitly as a by-product of exposure to, and communication in, the target language (Harley, 1998). However, evidence from immersion classrooms has suggested that

[11] A teaching technique for linking sentences *horizontally* ... with connectives (e.g. conjunctions). It can also cover sentence-embedding and other techniques for expanding and complicating the structure of sentences (p. iii).

even extensive exposure to the target language is not necessarily sufficient to promote grammatical accuracy in production (Spada, 1997). The focus in immersion classrooms tends to be overwhelmingly on meaning and whilst making considerable improvements in their communicative ability in the target language, learners can often overlook those grammatical features, which are not as necessary for successful communication (Harley, 1998; Long & Robinson, 1998; VanPatten, 2004a). It has therefore been argued that a certain amount of explicit form-focussed instruction integrated within a communicative teaching approach could be of benefit for young learners (Harley, 1998; Spada, 1997; J. White, 2008). By increasing their saliency learners could be encouraged to attend to key grammatical features within the input (Schmidt, 1994); as an instructional principle this could apply “equally well to the L2 learning of children as to adults” (Harley, 1998, p. 157).

Young L2 learners can be divided into three separate groups: early childhood (age 2 to 7), middle childhood (age 8 to 11) and older childhood (age 12+) (Philp et al., 2008, p. 5). Language learning in early childhood is characterized by a reliance on implicit learning (Muñoz, 2007). In contrast, in middle childhood a child’s L1 becomes more highly developed and increases in grammatical complexity (Philp et al., 2008). In addition children at this age become more logical in their thinking and develop greater metalinguistic awareness. This development continues into older childhood as learners begin to possess a greater capacity for abstract thought (Philp et al., 2008). It has therefore been argued that instruction which allows middle and older learners to make use of their developing language analytic skills and problem solving abilities, in conjunction with more implicit learning strategies, may result in more efficient and rapid progress in the acquisition of certain grammatical forms (J. White, 2008).

A number of studies have explored the role of explicit knowledge in language learning by young learners and have demonstrated the potential effectiveness of explicit grammar instruction for this age group (Bouffard & Sarkar, 2008; Hanan, 2011; Harley, 1998; J. White, 2008; J. White & Ranta, 2002; L. White, Spada, Lightbown, & Ranta, 1991). Harley (1998), for example, investigated the effect of focus on form instruction on learners’ (seven to eight year old French immersion pupils) proficiency in grammatical gender in French. The experimental intervention took the form of classroom games, which required the learners to pay attention to gender distinctions: for instance naming objects using the correct masculine (*un*) or feminine article (*une*); or performing an action when they heard a masculine (e.g. touch toes) or feminine noun (e.g. hands on head) (Harley, 1998, p. 163). Harley (1998) found that the learners improved in their ability to discriminate between masculine and feminine articles as well as demonstrating greater accuracy in their productive use of the correct gender article with familiar nouns. However the learners were not able to use their knowledge of noun endings in order to predict the gender of novel nouns, suggesting that the instruction had resulted in item rather than system learning. Harley (1998) argued that this was likely due to the volume of new vocabulary that was introduced in the instruction sessions. Consequently the learners may have been pre-occupied with the meaning of the novel vocabulary items and therefore unable to attend to the relevant grammatical features (Harley, 1998; VanPatten, 2004a). Despite this result, Harley's study demonstrated that employing focus-on-form tasks in the L2 classroom is feasible and can be helpful for young learners (Harley, 1998).

Similarly L. White, Spada, Lightbrown, and Ranta (1991) found that input enhancement activities were effective in promoting the syntactic accuracy (in L2 English question formation) of young francophone L2 English learners (aged 10 to 11). The instructed groups completed input enhancement activities designed to focus their attention on the placement of subjects, auxiliaries, and question words in English questions, and received explicit information and corrective feedback. The learners in the instructed group significantly outperformed the control group on a GJT and a sentence-level written production task, as well as more spontaneous oral communication task. In addition this improvement was still evident at delayed post-test. L. White et al (1991) therefore argued that these findings provided evidence to support the claim that input enhancement can bring about genuine changes in learners' interlanguage system.

In line with these findings, J. White and Ranta (2002) found that the provision of metalinguistic information regarding possessive determiners in English, coupled with contrastive L1/L2 information (Rule group), led to an improvement in the learners' use of the target feature as measured on an oral picture description task. Additionally the Rule group were found to outperform the Comparison group (who had received no explicit instruction relating to possessive determiner use) on a metalinguistic ability task (J. White & Ranta, 2002). These findings demonstrated that not only did the provision of metalinguistic information improve the learners' knowledge "about" the target feature, but also their knowledge of how to "use" the forms in oral communication (J. White, 2008; J. White & Ranta, 2002). It should be noted that a number of pupils from both the Rule (at pre-test) and Comparison groups (at post-test) had acquired the target grammatical rule without instruction. White and Ranta (2002) attributed this finding to individual learner differences in language analytic ability; those learners with high analytic ability were able to induce the target grammatical rule simply from previous exposure to target language input. For those learners with lower analytic ability, however, such 'rule-inducement' was not possible. Therefore it has been argued that form-focussing instruction could be beneficial for such learners in drawing their attention to the target feature within the input (Skehan, 1998).

Notably research evidence has also demonstrated the need for instructional packages to be *explicit enough* to enable learners to benefit. J. White (1998), for instance, found no difference between groups that had received typographically enhanced (bolding, underlining, italics) versus unenhanced input, in the learners' use of possessive determiners in English. J. White (1998) suggested that this finding of no difference might have been due to the nature of the input enhancement, which may not have sufficiently increased the salience of the target feature. Additionally given that none of the learners received explicit information relating to the target feature, the enhanced input may in fact have been more similar to *unenhanced* input, which, in the case of the learners in this study, appeared not to have been sufficient for learning. Therefore the learners may have benefitted from a type of instruction which more explicitly focussed their attention on the target feature. Moreover, whilst input flooding and typographical enhancement might be effective with early childhood learners who are solely dependent on implicit learning mechanisms, for middle and older childhood learners, more explicit instruction which allows them to make use of their developing language analytic skills may result in more efficient and rapid progress in the acquisition of certain grammatical forms (as argued by Philp et al., 2008; J. White, 2008). Finally, as in J. White and Ranta's (2002) study, there were a small number of learners across the different experimental groups who were able to demonstrate at least partial control of the target feature without any explicit instruction (J. White, 1998). Therefore it is important to note the role of individual differences, such as level of language analytic ability, which may mediate the extent to which individual learners' can benefit from more explicit instruction (Skehan, 1998; J. White, 1998, 2008; J. White & Ranta, 2002) (section 2.4).

Research evidence has also revealed that children as young as eight are able to develop metalinguistic awareness of their emerging L2 system (Bouffard & Sarkar, 2008). Metalinguistic awareness can be defined as the "ability to look at language as an object" (J. White & Ranta, 2002, p. 261) and further can serve a pedagogical purpose in pushing learners to conceptualise their linguistic organisation; "the learner interprets language structure and the required grammatical operations, co-constructs hypotheses and consolidates already-acquired forms" (Bouffard & Sarkar, 2008, p. 6). Swain and Lapkin refer to such dialogue as "student enactment of mental processes" (p. 329). Such techniques have been found to be beneficial in promoting young learners' metalinguistic awareness, for example, through noticing a problematic target grammatical form and how it is different from their own interlanguage (Swain, 1998). Notably however, further research is needed in order to investigate the extent to which such dialogue impacts learners' L2 use (i.e. comprehension, production).

In a study by Bouffard and Sarkar (2008) participants (aged eight to nine) were shown recordings of their performance on communicative activities and encouraged to discuss and analyse any errors which they observed. Bouffard and Sarkar (2008) found that the learners' ability to discuss errors improved considerably over the three month study. Further, the learners were able to use their analysis as a tool to improve their language awareness, for example of the link between their L1 (English) and their L2 (French) use. Moreover the findings demonstrated that it is possible to teach younger learners how to “draw on their grammatical knowledge to build their developing L2” (Bouffard & Sarkar, 2008, p. 21). As one example, over the course of the study the learners were able to co-construct a basic understanding of verb tenses in French. Such findings highlight that learners as young as eight years old are able to successfully attend to form and explore languages as “dynamic systems” provided they are taught how to (Bouffard & Sarkar, 2008). Similarly Hanan (2011) investigated the metalinguistic ability of young English learners of L2 German using a one-to-one oral task in which the learners were asked to identify and discuss key grammatical features of the target language. The learners demonstrated the ability to discuss and put forth hypotheses regarding the L2 grammar, through drawing on both their L1 and L2 explicit knowledge. This finding is in line with that of Bouffard and Sarkar (2008) who highlight the importance of learners having an awareness of both the L1 and L2, and argue that “the process of learning an L2 might be embedded in an understanding of how the L1 system works” (p. 21).

The research evidence presented above has demonstrated that explicit knowledge and instruction may have a beneficial role in supporting young learners’ second language learning (Bouffard & Sarkar, 2008; Harley, 1998; J. White & Ranta, 2002; L. White et al., 1991), provided that instruction is age-appropriate and tailored to suit the learners’ “cognitive and linguistic readiness for form-focussed instruction” (J. White, 2008, p. 194). However it is important to note that many of the studies investigating the effectiveness of explicit grammar instruction for young learners have been carried out with children learning within immersion classrooms (e.g. Bouffard & Sarkar, 2008; Harley, 1998; J. White & Ranta, 2002). In such an environment learners receive extensive amounts of target language input; therefore it could be argued that the form-focussing activities which were utilised in the above studies enabled the learners to explicitly notice the target grammatical feature, which was then, perhaps implicitly, reinforced and consolidated into the learners interlanguage through extensive exposure to instances of the target feature within the classroom input (N. Ellis, 2002; Schmidt, 1990). Notably the target features, question formation, possessive determiners *his* and *hers*, and gender encoded in articles, are likely to occur frequently within the classroom context. However, within the standard foreign language classroom, such as in the UK, learners are very often only exposed to input from one non-native speaker (the teacher), for limited amounts of time per week (Muñoz, 2006, 2008b). Consequently further research is needed in order to determine whether explicit grammar instruction can be useful in developing young learners’ understanding and *use* of key grammatical features, when learning and instruction occurs within the input-poor foreign language classroom context (RQ 1).

2.2 Role of explicit knowledge in language acquisition

2.2.1 Learning-acquisition / explicit-implicit dichotomy

A vast amount of research has been carried out investigating the effect of explicit grammar instruction in SLA and in general a positive effect of instruction has been observed (Norris & Ortega, 2000; Spada & Tomita, 2010). Nevertheless there is by no means unanimity within the SLA research community as to the role of instruction or the type of knowledge which instruction makes available to the learner and differing views stem from disagreement as to the role of explicit knowledge in SLA.

2.2.2 Explicit versus implicit knowledge

Explicit and implicit knowledge can be defined in terms of awareness and automaticity. Explicit knowledge, also referred to as learned knowledge (Krashen, 1982) or declarative knowledge (DeKeyser & Criado, 2012), is conscious, verbalisable knowledge about the structure of language, which tends to be utilised in controlled processing, for example when a learner encounters difficulty in their L2 use (R. Ellis, 2006). R. Ellis (2006) identifies two facets of explicit knowledge: analysed knowledge which entails a "conscious awareness of how a structural feature works"; and metalinguistic explanation, which refers to "knowledge of grammatical metalanguage and the ability to understand explanations of rules" (p. 95). In contrast a learner's implicit knowledge about language, otherwise termed acquired (Krashen, 1982) or procedural knowledge (DeKeyser & Criado, 2012), is the knowledge which they are not aware of possessing. It is the unconscious, procedural knowledge, which is employed automatically during spontaneous language use (R. Ellis, 2006). A learner's linguistic competence (i.e. innate linguistic knowledge) is comprised of their *implicit* knowledge about language, and has been described as being "intuitive and tacit, rather than conscious and explicit" in nature (R. Ellis, 2005, p. 143). However, opinion differs as to whether explicit knowledge can have any impact on a learner's linguistic competence (R. Ellis, 2005). Consequently SLA research has been concerned with whether explicit grammar instruction can lead to the acquisition of target grammatical features, or whether it simply results in explicit knowledge of the feature in question (Norris & Ortega, 2000).

2.2.3 Non-interface hypothesis

Proponents of the non-interface hypothesis argue that explicit and implicit knowledge are the product of two distinct language acquisition mechanisms (Hulstijn, 2002; Krashen, 1982).

Krashen (1982) draws a clear distinction between acquisition (implicit learning) and learning (explicit learning), arguing that only acquisition leads to the development of "acquired competence", or implicit knowledge (p. 10). Through acquisition, learners are not aware of the rules they have acquired, rather they develop intuitions about the correctness of the language (Krashen, 1982). Learning on the other hand results in the development of "knowledge about a language…grammar and rules", i.e. explicit knowledge, and an awareness of grammatical rules and the ability to talk about them (Krashen, 1982). The non-interface position posits that it is not possible for explicit knowledge to become implicit knowledge, therefore knowledge gained via explicit learning cannot become part of the learners' linguistic competence and cannot be made available for spontaneous language use (R. Ellis, 2005). Nevertheless it is acknowledged that it is possible for implicit knowledge to transform into explicit knowledge, via conscious reflection on output which was first produced by a learners' implicit knowledge (Bialystok, 1994). Additionally the Monitor Hypothesis (Krashen, 1982) claims that explicit knowledge has only one function as a "monitor" to make changes to utterances produced via the acquired system. Therefore conscious learning (e.g. via explicit grammar instruction) has only a limited role to play in enabling the monitoring of second language performance. Research such as that by Paradis (1994) lends support to this view by proposing that the implicit and explicit memories are neurologically distinct. In line with Krashen's standpoint, Paradis (1994) argues that explicit, learned knowledge cannot become procedural, implicit knowledge, and further than explicit knowledge cannot be made available for use as part of the automatic production process.

Krashen's (1982) acquisition-learning hypothesis posits that all language learners, whether child or adult learning their native or a second language, are able to acquire language via access to the same natural language acquisition device. Consequently supporters of the non-interface hypothesis maintain that language instruction should focus on providing learners with rich and varied *comprehensible* input and opportunities to practice using the language in meaningful and spontaneous interactions (L. White et al., 1991). In contrast instruction focussing on the teaching of grammatical form is considered superfluous and to a certain extent counter-productive (L. White et al., 1991). Nevertheless it is important to note that some supporters of the non-interface hypothesis do see a value in explicit teaching, arguing that whilst explicit knowledge cannot become implicit knowledge, "explicit learning and practice often form efficient ways of mastering an L2 by creating opportunities for implicit learning" (Hulstijn, 2002, p. 193).

2.2.4 Strong interface hypothesis

In opposition to the non-interface hypothesis, the strong interface hypothesis posits that knowledge learnt explicitly *can* become implicit knowledge, which is available for spontaneous communication, via automatization through extensive practice (Anderson, 1992; Bialystok, 1994; DeKeyser & Criado, 2012; R. Ellis, 2005; Norris & Ortega, 2000). DeKeyser (1998) emphasises the necessity for that practice to be communicative, rather than mechanical, in nature, arguing that in order for declarative knowledge to be proceduralised the learner needs to engage in activities which require the use of the relevant declarative knowledge in order to convey meaning. Further the strong interface position holds that explicit knowledge can become implicit if the learner loses awareness of a particular structure over time, and that the opposite is also true; implicit knowledge can become explicit if the learner develops awareness of a structure when applying it to a new context or explaining it to a third party (DeKeyser, 2003)

Theories regarding the automatization of explicit, or declarative, knowledge have been proposed such as 'Skill Acquisition Theory' (SAT) (DeKeyser, 2007). The fundamental premise of SAT is that language acquisition consists of "a series of sequenced stages, from initial representation of knowledge to highly skilled behaviour" (DeKeyser & Criado, 2012, p. 1). SAT identifies three types of language knowledge (DeKeyser, 2007; DeKeyser & Criado, 2012). *Declarative* knowledge is equivalent to explicit knowledge, and refers to knowledge of facts about language such as specific grammatical rules. *Procedural* knowledge can be described as "knowing how" to perform particular behaviours and is associated with implicit, unconscious knowledge. Finally *automatized* knowledge is the result of the restructuring of proceduralised knowledge which occurs during meaningful practice and is the source of rapid and fluent language use. Similarly Anderson's Adaptive Control of Thought (ACT) model identifies declarative knowledge as knowledge of facts, whereas procedural knowledge can be defined as the knowledge of how to perform particular cognitive activities (Anderson, 2005).

Theories, such as SAT and ACT, posit that explicit, declarative knowledge can gradually become specialized, proceduralised knowledge, which is in turn automatized, as a result of meaningful practice (DeKeyser, 1998, 2007). Automatization can be characterized as a continuum along which explicit, declarative, knowledge develops into implicit, proceduralised, knowledge, resulting in a decline in reaction times and error rates as well as a reduction in interference from and with simultaneously performed tasks (DeKeyser, 2007).

A position such as the strong interface position, therefore, promotes the use of explicit grammar instruction as a means of providing explicit information coupled with opportunities for practice, leading to the eventual automatization of the explicit knowledge (R. Ellis, 2006). Nevertheless it is important to note that criticisms have been raised of such a position, regarding the question of whether declarative knowledge learnt through explicit instruction can become "a form of procedural knowledge that is accessible in the same way as implicitly acquired knowledge" (DeKeyser, 2003, p. 328). One's interpretation of this issue depends upon how acquired, or implicit, knowledge is defined. First, if implicit knowledge is defined by a lack of awareness (e.g. Rebuschat, 2013; Rebuschat & Williams, 2012), then proceduralised explicit knowledge cannot be described as "implicit" if the learner retains a level of awareness of the grammatical rule in question (DeKeyser, 2003). Additionally Hulstijn (2002) argues that practice will speed up the "execution of algorithmic rules to a limited extent" (p. 211); however maintains that there is a fundamental difference between automated explicit knowledge and implicit knowledge in terms of the extent to which it can be accessed during automatic processing. Nevertheless, DeKeyser (2003) argues that in terms of automaticity, if the procedural knowledge (or automatized declarative knowledge) developed following instruction is accessible with the same level of automaticity (for example indicated by a reduction in reaction time and error rate as in DeKeyser, 1997) as implicit knowledge then it can be considered to be functionally equivalent to implicit, acquired knowledge.

2.2.5 Weak interface hypothesis

The weak interface hypothesis offers a distinct standpoint from both the non- and strong interface positions. Whilst implicit learning is considered the "default" acquisitional process (N. Ellis, 2005; J. White & Ranta, 2002), advocates of the weak interface hypothesis argue that explicit knowledge and instruction do have an important role. There are several formulations of the weak interface hypothesis. Firstly it is argued that practice can lead to explicit knowledge converting into implicit knowledge on the proviso that the learner is developmentally ready to acquire the relevant linguistic form (R. Ellis, 2005). Such a position is in accordance with the developmental sequences set out in Pienemann's (1989) teachability hypothesis.

Secondly an alternative formulation of the weak-interface hypothesis posits that explicit knowledge can play a role in *facilitating* the development of implicit knowledge, through drawing the learner's attention to specific features in the input (R. Ellis, 2005; Schmidt, 1990; VanPatten, 2002). N. Ellis (2005) argues that although Krashen (1982) and Paradis (1994) were correct in so far as to say that explicit and implicit knowledge are distinct, he contends that there is an interaction between the two knowledge types. N. Ellis (2005) proposes that the interface is dynamic and "happens transiently during conscious processing" (p. 305), however it can have a lasting effect on the learner's implicit knowledge. Explicit learning provides the learner with the initial explicit representation of the form-meaning association of the target feature, which is subsequently integrated into the developing system through implicit learning during input processing (N. Ellis, 2005). Similarly the Implicit Tallying Hypothesis posits that following the initial explicit registration of the target form in the input, the form and "its associations will be tallied and implicitly catalogued" (N. Ellis, 2002, p. 174), during subsequent exposure to repeated instances of the target feature.

Proponents of the weak interface hypothesis therefore argue that explicit grammar instruction can facilitate the process of "noticing" by increasing the saliency of key grammatical forms. Noticing coupled with subsequent exposure to repeated instances of the target feature within the input, can result in the acquisition of the target grammatical feature (Schmidt, 1990). Additionally explicit knowledge can enable learners to perform the "cognitive comparison" between what is observed in the input and their own output (R. Ellis, 2006, p. 97), which is identified as a key acquisitional process, termed "noticing the gap" (Schmidt & Frota, 1986). It is also argued that explicit grammar instruction can serve to speed up acquisition by pushing learners to attend to problematic grammatical features within the input (N. Ellis, 2005).

2.2.6 Operationalising explicit and implicit knowledge

As can be seen above, debate is rife as to the relationship between explicit and implicit knowledge and the impact of this relationship on the acquisition of language. Whilst numerous studies have investigated whether there is any discernible value in teaching grammar explicitly (Norris & Ortega, 2000; Spada & Tomita, 2010), relatively little research has sought to test whether any such interface exists between the two knowledge types or whether it is possible for explicit knowledge to convert directly into implicit knowledge (R. Ellis, 2005, 2006). The main reason for the lack of empirical evidence to support or refute the interface hypotheses set out above concerns the difficulty of designing tasks, which provide pure and sensitive measures of implicit and explicit knowledge (DeKeyser, 2003; Rebuschat, 2013). Notably it is difficult to determine what type of knowledge learners are utilising when they are completing a given test or task (R. Ellis, 2006). Indeed R. Ellis (2002) highlights that all tests are likely to draw on both the learner's implicit and explicit knowledge to a certain extent.

Bearing in mind these methodological considerations, one aim of the present study is to contribute to the interface debate by investigating whether explicit grammar instruction can lead to the development of some knowledge bearing some of the characteristics of implicit, as well as explicit, knowledge of the target grammatical feature (RQ 2). As such it is important to establish how explicit and implicit knowledge will be operationalised.

As discussed above, explicit and implicit knowledge can be distinguished based on differences in levels of awareness and automaticity. Accordingly R. Ellis (2009b) identifies a number of criteria by which these two characteristics can be defined; pertinent to this study are *time available*, *focus of attention*, and *metalanguage*. Based on these criteria, implicit knowledge will be operationalised as the knowledge employed when learners are performing online, time-pressured tasks (*time available*) and are focussed on conveying meaning (*focus of attention*). In contrast, explicit knowledge will constitute the knowledge used when the learners are not under any time pressure (*time available*) and are focussed on form rather than meaning (*focus of attention*).

An additional component of the learner's explicit knowledge will be operationalised as their metalinguistic ability (*metalanguage*). Metalinguistic ability can be defined as the ability to look at language as an object, to focus on the structure rather than the meaning of a message (Ryan & Ledger, 1984, cited in J. White & Ranta, 2002). Metalinguistic ability can further be expressed as a continuum, characterised by the dimensions of explicitness and elaboration, and can include for example briefly attending to a particular linguistic form versus the detailed explanation of a grammatical rule using sophisticated metalinguistic information (Sharwood Smith, 1991). Notably it is important that metalinguistic ability is treated as a related, yet separate, component of explicit knowledge. Being consciously aware of, and able to use, a particular grammatical feature does not necessarily mean that the learner is able to articulate the relevant grammatical rule, and vice versa (Harley, 1998; J. White & Ranta, 2002).

The outcome measures to be utilised in the present study will be developed in accordance with the above operationalization of implicit and explicit knowledge (Section 3.6).

2.3 Explicit grammar instruction and SLA

Despite disagreement as to the role of explicit knowledge in SLA, numerous studies have sought to investigate the effectiveness of explicit grammar instruction in instructed learning, and in general a positive effect of instruction has been observed (Doughty & Williams, 1998a; Norris & Ortega, 2000; Spada, 1997; Spada & Tomita, 2010). Further, studies comparing instructed versus naturalistic learning have demonstrated that whilst all learners, regardless of language learning setting, tend to follow the same order of acquisition (Pienemann, 1989), instructed learners progress more quickly and achieve a higher level of proficiency than those learners learning within an immersion setting (R. Ellis, 2006). Norris and Ortega's (2000) comprehensive meta-analysis of effect of instruction research found that explicit instruction was more effective than implicit instruction and resulted in substantial, target oriented, durable gains, although arguably on fairly controlled measures (Truscott, 2004). Similarly Spada and Tomita (2010), in their meta-analysis of research studies investigating the interaction between different types of instruction and the complexity of grammatical features, observed more substantial gains resulting from explicit rather than implicit instruction for both simple and complex features. Consequently in recent years the focus of research has shifted from investigating whether or not explicit instruction may be useful, to determining which type of explicit grammar instruction is most effective in promoting learning within the instructed setting (Norris & Ortega, 2000) and on the precise measures that learning is observed on (section 2.3.5).

2.3.1 Focus on form, focus on formS, or form-focussed?

Explicit grammar instruction is an all-encompassing term which can refer to a host of instructional techniques, from traditional approaches with a core focus on form to more communicative-oriented approaches in which attention to form arises organically through activities which are primarily focussed on meaning (R. Ellis, 2001).

Long (1991) subdivided explicit grammar instruction into two approaches; 'focus on *form*' (FonF) and 'focus on *formS*' (FonFS). The emphasis in FonFS instruction is on teaching the formal elements of language in discrete and isolated chunks (Doughty & Williams, 1998a; Long & Robinson, 1998). The assumption underpinning FonFS approaches is that grammatical features should be taught one-by-one in a specified sequence, which is pre-determined based on linguistic complexity (Norris & Ortega, 2000). At the other extreme lies focus on meaning (FonM) instruction, which emphasises an implicit approach based on the provision of positive evidence and opportunities to practice communicating in the target language (Long & Robinson, 1998), to the exclusion of any focus on grammatical form (Doughty & Williams, 1998a).

In contrast, FonF instruction draws on key elements from the FonFS and FoM approaches. The fundamental premise of FonF techniques is that learners need to be focused on meaning first, before any subsequent attention to grammatical form can be effective (Doughty & Williams, 1998a). FonF tasks are designed according to the communicative needs of the learners, for example to provide practice using language appropriate to a job interview or making holiday reservations (Long & Robinson, 1998). Crucially there is no predetermined linguistic focus, rather attention to form arises incidentally during task completion, for example due to the learners encountering a problem in their comprehension and / or production of a particular feature (Long & Robinson, 1998). Notably this attention to form constitutes only an "occasional shift of attention to linguistic code features" (Long & Robinson, 1998, p. 23). Consequently Long and Robinson (1998) advocate the use of unobtrusive techniques, such as input flood, input enhancement, explicit negative feedback, as well as implicit negative feedback (e.g. some types of recasts) in order to increase the perceptual salience of the target feature whilst maintaining a focus on meaning. The aim of such techniques is to increase the likelihood of the learner *noticing* the problematic grammatical form in the input, which is subsequently more likely to be available for intake into the learner's developing interlanguage (Long & Robinson, 1998; Schmidt, 1990, 1994) (see section 2.3.2).

Arguably Long and Robinson's (1998) definition of FonF instruction lies at the more incidental end of the instruction spectrum (Doughty & Williams, 1998a). Consequently it can be argued that the type of instruction investigated in many studies claiming to employ FonF (e.g. Harley, 1998; VanPatten & Cadierno, 1993; J. White, 1998) may in fact be more akin to FonFS tasks if for example the tasks were designed with a specific formal linguistic focus or included the provision of explicit information about the target grammatical feature (Doughty & Williams, 1998a). Nevertheless it can also be argued the such instruction is compatible with certain elements of Long and Robinson's (1998) definition of FonF, since the tasks are often designed with a strong focus on meaning and are employed in reaction to the problems that learners have encountered in their language use rather than "imposed externally by a linguistic syllabus" (Doughty & Williams, 1998a, p. 5). Therefore Doughty and Williams (1998a) propose a less restrictive definition of FonF instruction, which incorporates both attention to form which arises reactively in response to learner difficulties during a meaning-based task (in line with Long and Robinson's definition) and tasks which are proactively designed in order to target a grammatical feature which has previously been observed to be problematic for learners.

A number of researchers have also defended the use of metalinguistic information in instructional tasks. Whilst Long and Robinson's (1998) FonF tasks would arguably not include a metalinguistic focus, a number of studies have found that the provision of metalinguistic information constituted a key element of the instructional package which contributed to the learners' success (e.g. Bouffard & Sarkar, 2008; Culman, Henry, & VanPatten, 2009; Harley, 1998; Swain, 1998) (see section 2.3.4.4).

In order to circumvent such issues in the defining and operationalising of FonF versus FonFS instruction, Spada (1997) employs the term *form-focussed instruction* to refer to any meaning-based instruction which incorporates *either* a spontaneous *or* a predetermined focus on language. Similarly R. Ellis (2001) defines form-focussed instruction as any "planned or incidental instructional activity that is intended to induce language learners to pay attention to linguistic form" (p. 1). The crucial difference between *form-focussed instruction* and Long's (1991) *focus-on-form* is the unequivocal requirement in FonF instruction that attention to form may only be employed spontaneously as the need arises in otherwise meaning-focussed activities (Spada, 1997). Nevertheless in her review of studies investigating the effects of form-focussed instruction, Spada (1997) concludes that explicit form-focussed instruction (i.e. instruction with a predetermined linguistic focus and / or including the use of metalinguistic information) can be beneficial within a communicatively-oriented classroom. Notably instruction involving the provision of metalinguistic information as well as corrective feedback was found to result in greater improvements than more implicit instruction without these elements (Spada, 1997). Further Norris and Ortega (2000) found comparative effects for FonF and FonFS instructions, providing further support for a less restrictive definition of meaning-focussed instruction which includes some kind of focus on more formal elements of language. Consequently with regard to the present study explicit grammar instruction will be operationalised in line with the principles of *form-focussed instruction*:

- the instruction will be designed in order to target a specific linguistic feature of the target language
- the target grammatical feature will be chosen based on a particular processing problem which L2 learners of the target language commonly encounter
- the instruction will retain a primary focus on meaning (at sentence-level)
- the instruction will include metalinguistic information related to the target grammatical feature

2.3.2 Role of input

Language learning has been described as a three-way interaction between the input, the learner, and the learning context (MacWhinney, 2001). Exposure to, and the processing of, target language input have therefore been identified as playing a key role in the development of a learner's internal grammar (Long, 1990; VanPatten, 1996). The nature and extent of target language input, to which language learners are exposed, are necessarily fundamental contributors to success in acquiring grammar in a second or foreign language. Krashen (1981, 1982) states that language acquisition occurs via 'natural' processes as a result of exposure to "comprehensible input" and in his input hypothesis claims:

> "a necessary (but not sufficient) condition to move from stage i to stage i + 1
> is that the acquirer understand input that contains i + 1, where "understand"
> means that the acquirer is focussed on the meaning and not the form
> of the message" (Krashen, 1982, p. 21)

Comprehensible input can therefore be defined as target language input which is 'one step ahead' of the learner's current level of competence. With the help of contextual and extra-linguistic clues the learner is able to understand and ultimately acquire the new input (Krashen, 1981, 1982). In addition, contrary to traditional grammar teaching approaches, Krashen argues that target language grammar is acquired as a by-product of the learner "going for meaning". From this standpoint it can therefore be argued that input, provided it is appropriate to the level and ability of the learner, is the necessary and sufficient condition for acquisition to occur. Consequently Krashen (1981) argues that there is no role for explicit grammar instruction in terms of promoting learners' acquisition of the target language, rather "the major function of the second language classroom is to provide intake for acquisition" (p. 101) (see also section 2.2.3).

Nevertheless it has been argued that exposure to comprehensible input alone is not necessarily sufficient to facilitate acquisition (DeKeyser, 1998; Reinders & Ellis, 2009; Spada, 1997; VanPatten, 2004a). Evidence from research carried out within an immersion setting has demonstrated that learners can often fail to reach native-like competency in certain features of an L2, even after extensive exposure to the target language (Doughty & Williams, 1998b; Harley, 1998; Loewen, Erlam, & Ellis, 2009; Spada, 1997). Such difficulties can be attributed to the fact that some grammatical features may be less salient or semantically redundant within the input (Reinders & Ellis, 2009; VanPatten, 2004a). For instance a study by Loewen et al. (2009) demonstrated that even after intensive exposure to input containing exemplars of the third person -s, the learners failed to acquire the target feature; thereby providing evidence in support of the above claims. Notably such claims can also be extrapolated to the foreign language classroom setting, in which exposure to the target language is limited to such an extent that learners are unlikely to receive sufficient levels of comprehensible input in order for acquisition to occur. It has been proposed that whilst positive evidence alone is sufficient to result in acquisition for many grammatical features, explicit form-focussed instruction can play a role in helping learners to notice those cues which are, for example, less salient, less frequent, less like the L1, or more communicatively redundant (Doughty & Williams, 1998b; Reinders & Ellis, 2009; Sharwood Smith, 1991; Spada, 1997; VanPatten, 2002, 2004a).

2.3.2.1 Noticing, attention, and awareness in SLA

Based on the observation that the provision of input alone may not be sufficient to result in acquisition, a key concern of research into form-focussed instruction has been to determine in what ways input can be enhanced to promote attention to and noticing of grammatical features in a useful way for learning (R. Ellis, 2001). This question is motivated to a large extent by theories relating to the role of attention and awareness in acquisition (Doughty, 2003).

A key theoretical position was put forward by Schmidt (1990, 1995, 2001) who hypothesised that language learning requires a level of conscious processing, which is underpinned by "what learners pay attention to and notice in the input" (Schmidt, 2001, p. 3). Notably it is this assertion which stands Schmidt's hypothesis in opposition to Krashen's claims that acquisition can only occur unconsciously. There are three ways in which the term "unconscious" can be interpreted within the context of language learning; learning without intention, learning without explicit knowledge, and learning without awareness (Robinson, 2003; Schmidt, 1990). It is this last interpretation which Schmidt (1990) takes issue with, arguing that learning has to involve awareness; learners must consciously attend to, and "notice", features in the input in order for learning to take place. Such a position has come to be known as the Noticing Hypothesis, in which Schmidt (1990) states that "noticing is the necessary and sufficient condition for converting input to intake" (p. 131). It is important to note that in this context intake refers to the subset of input which is made available for further processing and subsequently acquisition (Reinders & Ellis, 2009; Schmidt, 1995; Truscott, 1998).

Noticing can be defined as the conscious registration of the presence of a grammatical feature in the input (Robinson, 1995; Schmidt, 2001). According to the Noticing Hypothesis conscious awareness *at the level of noticing* has to occur in order for a grammatical feature to be acquired. Notably this can be distinguished from awareness *at the level of understanding*, which entails metalinguistic awareness of the rules which govern the use of a particular language feature (Schmidt, 1995). Awareness at the level of understanding is not required in order for learning to take place, but can play a facilitative role in learning (Robinson, 2003; Schmidt, 1990). Schmidt's claims are grounded in observations based on his own language development as an L2 learner of Portuguese. Schmidt and Frota (1986) found a clear relationship between Schmidt's language use and his personal reflections on the language input he was exposed to in class and from native speakers. Only once features were consciously noticed, i.e. commented on in Schmidt's language diary, did they appear in his language use. Such evidence is taken to support the relationship between noticing and acquisition as proposed in the Noticing Hypothesis.

Nevertheless the role of attention and awareness in acquisition remains contested. Contrary to Schmidt's claim that acquisition requires attention in the form of conscious awareness (at the level of noticing), Tomlin and Villa (1994) maintain that "detection" is the necessary and sufficient level of attention required for acquisition. Detection constitutes the "cognitive registration of some stimuli" (Tomlin & Villa, 1994, p. 190). Crucially detection can be disassociated from awareness, awareness in this sense being defined as the "subjective experience of some cognitive content or external stimulus" (Tomlin & Villa, 1994, p. 193). Semantic priming studies (e.g. Marcel 1983) have been taken as evidence to support the claim that detection can occur without awareness. For example, Marcel (1983) demonstrated, in a speed reading task, that participants exhibited faster reading times when target words were primed by semantically related words (e.g. doctor – nurse) than by non-related words (e.g. doctor – balloon), even when participants were not aware of having read the priming word. Notably however such studies do not necessarily provide evidence of the learning of novel input, but rather that existing knowledge can be activated automatically, without awareness; a finding which is not incompatible with the Noticing Hypothesis (Robinson, 2003; Schmidt, 1995, 2001).

As an alternative explanation of the role of attention in SLA, compatible with the position set forth in the Noticing Hypothesis, Robinson (2003) proposes that acquisition occurs as a result of detection coupled with rehearsal in short term memory. Specifically in order for newly detected information to be learned, “the information must enter focal attention and so short-term memory, where rehearsal processes operate prior to encoding in long-term memory” (p. 654). In this sense then detection can be defined as “recognition outside of awareness in passive short-term memory” (Robinson, 2003, p. 655). This recognition coupled with the necessary rehearsal processes subsequently results in noticing and a higher level of awareness. From this position, awareness can therefore be said to be a product of these processes but is not necessary in order to explain their occurrence (Schmidt, 1995, p. 28).

Given the different theoretical standpoints proposed and the lack of consensus within the literature, a substantial amount of research (e.g. Grey, Rebuschat, & Williams, 2014; Hama & Leow, 2010; Leow, 2015; Leung & Williams, 2011; Robinson, 1997; Rosa & Leow, 2004a; Williams, 2005) has focussed on exploring the hypothesised role of attention and awareness in SLA with mixed results. Rosa and Leow (2004) for example compared the role of awareness for different learning conditions in which learners completed computerized problem-solving tasks (targeting the past conditional in Spanish), which varied in their degree of explicitness; explicitness was controlled via the provision, or withholding, of explicit pre-task information and the provision of explicit versus implicit feedback. The participants’ awareness of the target feature was gauged by means of both online (think-aloud) and offline (post-exposure awareness questionnaire) measures. Rosa and Leow (2004) found that higher levels of awareness (i.e. at the level of understanding) were more prevalent for learners in the explicit conditions and further were associated with the most substantial learning gains as measured by controlled tests of recognition and production. Notably awareness at the level of noticing was also found to result in positive learning gains, although not to the same extent as awareness at the level of understanding.

Similarly Robinson (1997), in his study comparing learning under Implicit and Incidental versus Rule-search and Instructed conditions, utilised a “debriefing questionnaire” in order to investigate learners’ awareness of easy versus hard second language rules. Learning was assessed via a grammaticality judgement task, and in line with Rosa and Leow’s findings, Robinson (1997) observed that only the highest level of awareness, termed “ability to verbalise rules” (p.78) bore a significant relationship to the learning outcomes for both simple and hard rules, for all of the groups. Notably, however, awareness at the level of noticing was not found to improve the learners’ performance, a finding which is contrary to that of Rosa and Leow

(2004a). The discrepancy in findings between the two studies may be a result of the fact that different measures of learning were utilised, as well as different methods for eliciting awareness data (Rosa & Leow, 2004a). Further the nature of the input the learners were exposed to differed between the two studies, with Rosa and Leow (2004a) following a task-based approach, which actively engaged the learners in problem-solving tasks, whereas the participants across Robinson's (1997) four conditions were simply required to respond to discrete questions after reading each sentence (for example related to the location of certain words in the input or whether they had noticed a relevant rule).

In contrast to Rosa and Leow's (2004a) and Robinson's (1997) findings, Williams (2005) found that learning without awareness (at the level of noticing and / or understanding) can occur. Participants were given instruction on a miniature noun class system (distance form meaning connections), however were not told that the correct choice of determiner was dependent upon the animacy of the accompanying noun. Williams (2005) found that most participants remained unaware of this form-meaning connection (FMC), as measured on retrospective post-exposure reports. Nevertheless when tested the participants demonstrated above chance reliability in choosing the determiner with the correct animacy correspondence for each of the target nouns. Similarly Leung and Williams (2011) reported evidence of learning without awareness of contextually derived FMCs[12]. Leung and Williams (2011) observed an increase in reaction times for critical conditions in which the target FMC had been violated, even for those learners (80%) who had reported being unaware of the connection following testing.

One important consideration of such studies is the methodological issue of how awareness is measured (Hama & Leow, 2010; Leow, 2001, 2015; Robinson, 2003). Indeed Truscott (1998) put forward a fundamental criticism of Schmidt's Noticing Hypothesis on the basis that it is unfalsifiable since it is nigh on impossible to precisely and definitively measure whether or not awareness has occurred. Nevertheless a number of studies have attempted to measure awareness, however, as illustrated above, have produced differing results. Disagreement as to the interpretation of findings stems in part from the fact that some studies have employed online measures, whereas others have employed offline measures of awareness. In order to address such methodological issues, Hama and Leow (2010) carried out a replication of William's (2005) study and employed an online measure of awareness (think-aloud), as well as the offline task (post-exposure report) utilised in the Williams' study. Contrary to Williams (2005), however, Hama and Leow (2010) found that for those learners who were "unaware" of the target feature at the stage of encoding, learning did not occur. But

[12] Articles *gi* and *ul* referred to agent, *ro* and *ne* to patients.

again other methodological differences (e.g. number of answer options in the assessment task) could have contributed to the difference in findings.

The theoretical perspectives and empirical evidence presented above demonstrate that as yet there is no consensus as to whether awareness at the level of noticing, as specified in the Noticing Hypothesis, is required for learning. Put simply, it is not yet clear from research findings whether learning can happen when learners are *unaware*. It is also important to consider that such studies have tended to employ very specific measures of learning (e.g. reaction times). Therefore it remains to be seen whether learning would also be evident on more spontaneous production measures (Leung & Williams, 2011). Nevertheless in all of the studies discussed above awareness was consistently associated with learning and it can therefore be argued that awareness can be beneficial in enhancing learners' input processing (Rosa & Leow, 2004a). The next section addresses constraints on such noticing and how it might be enhanced.

2.3.2.2 Barriers to noticing

As mentioned above, previous research has clearly demonstrated that even in naturalistic or immersion settings second and foreign language learners often fail to acquire certain grammatical features in the input (Doughty & Williams, 1998b; Harley, 1998; Loewen et al., 2009; Spada, 1997; VanPatten, 2002). Consequently it can be argued that positive evidence alone, i.e. input, is not necessarily sufficient to guarantee that learners attend to (whether with or without awareness) and subsequently acquire all grammatical features in a second or foreign language (Spada, 1997). Before exploring the various instructional options which have been proposed in order to address this issue, it is important to consider why some grammatical features appear to be more easily acquired than others.

i) Low saliency and redundancy

Firstly it has been proposed that saliency plays a key role in mediating whether or not a feature is attended to, with characteristics such as frequency and semantic redundancy affecting the saliency of a given grammatical feature within the available input (N. Ellis, 2006; R. Ellis, 2006; Reinders & Ellis, 2009; Schmidt, 2001; Sharwood Smith, 1991; Spada, 1997; VanPatten, 2004a); the lower the saliency of a grammatical feature, the less likely it is that that feature will be attended to in the input and subsequently acquired. Additionally, Goldschneider and DeKeyser (2001) observed that the order of acquisition of L2 English grammatical morphemes can to a certain extent be determined by a combination of factors: perceptual salience, semantic complexity, morpho-phonological regularity, syntactic category, and frequency. Further redundancy has been highlighted as a key factor contributing to low saliency, which therefore plays a significant role in inhibiting attention to, and noticing of, certain grammatical features

(N. Ellis, 2006; VanPatten, 2004a). For instance, very often sentences will tend to include lexical items, such as temporal adverbs, which convey the same meaning as grammatical inflections for past tense, as in the following example:

Gestern sind wir ins Kino gegangen.

Yesterday we went to the cinema.

Both the auxiliary (*sind*) and past participle (*gegangen*) are redundant, as the sentence is fronted by the adverb *gestern* (yesterday), which conveys that the event happened in the past (N. Ellis, 2006). In line with this observation, VanPatten (2004a) proposes that "learners will tend to rely on lexical items as opposed to grammatical form to get meaning when both encode the same semantic information" (p. 9), a proposal which has been termed the Lexical Preference Principle (see section 2.3.3). It has therefore been proposed that instruction can serve to increase the saliency of key features within the input, which it can be argued, in line with the Noticing Hypothesis, will increase the likelihood of their being attended to, or noticed, and subsequently acquired (DeKeyser, 2003; Doughty & Williams, 1998b; R. Ellis, 2006; Schmidt, 1995, 2001; Svalberg, 2012).

ii) Limited attention

Whilst it is generally agreed that attention facilitates learning, it is also important to note that it is considered by some to be a limited resource (N. Ellis, 2006; Svalberg, 2012; VanPatten, 2004a). Findings from studies conducted within the divided attention paradigm have demonstrated lower levels of learning (e.g. of sequences of numbers) in dual-task conditions compared to single-task conditions (Schmidt, 1995). With regard to extrapolating this finding to SLA, some have argued that, when processing target language input received through communicative interaction, learners are constantly engaged in a "dual-task", attending to multiple levels of meaning as well to linguistic form (Schmidt, 1995; VanPatten, 2004a). Further it is generally assumed that the learning of grammatical features will require a certain level of focal attention (Robinson, 2003; Schmidt, 1995; Tomlin & Villa, 1994), which may vary due to the inherent complexity of a particular feature (Schmidt, 1995). On this basis it has been suggested that 'simple' structures may require lower levels of focal attention and will therefore be more easily acquired in uninstructed settings, whereas more 'complex' syntax will require a greater level of focal attention and therefore may benefit from decontextualisation through focused instruction (Schmidt, 1995).

In line with these observations VanPatten (2004a) argues that in order for learners to be able to attend to and process grammatical forms in the input, the processing of overall sentential meaning must not drain the available (limited) processing resources (Availability of Resources principle, Appendix 1). If the learner is preoccupied with processing novel lexical items in the input, there will not be sufficient processing resources available in order for them to simultaneously process target grammatical features (see section 2.3.3).

However, it is also important to note that some have taken issue with the proposal that attention is a limited resource, arguing that accounts of limited attention such as that put forward by VanPatten (1996, 2004a) fail to provide an explanation as to how and why attention is limited (DeKeyser, Salaberry, Robinson, & Harrington, 2002). In contrast it has been argued that attentional resource capacity is in fact unlimited (DeKeyser et al., 2002; Robinson, 2003). From such a perspective, in order to account for why certain grammatical features may not be acquired, Robinson (2003) argues that, whilst focal attention facilitates the selection of input to be converted to intake, it may also act to inhibit the detection of certain elements within the input, in order to maintain "continuity of action" and prevent interference (p. 638). Further DeKeyser et al. (2002) argue, contrary to VanPatten (2004) and Schmidt (1995) that attending to grammatical form as well as meaning within language input does not constitute a dual-task, rather it is a single task drawing on the same "verbal encoding resource pool" (p. 809).

iii) L1 transfer

Notably, and perhaps somewhat obviously, L2 learners are distinct from L1 learners, due to the fact that they come to the task of acquiring an L2 having already achieved fluency in their first language (Bley-Vroman, 1990; N. Ellis, 2010; MacWhinney, 2005; Philp et al., 2008). This knowledge of a first language has been shown to impact and lead to difficulty in learners' L2 acquisition, due to, for instance, misleading similarities between certain grammatical features in the L1 and L2 (Spada, Lightbown, & White, 2005). Further it has also been proposed that factors such as redundancy are likely to be more influential in L2 rather than L1 acquisition due to the L2 learner's previous language experience (N. Ellis, 2006). Based on their L1 experience, L2 learners will have expectations about the L2 language input they are exposed to, and will therefore be aware, for example, that temporal adverbs tend to co-occur with particular tenses, thereby increasing their saliency in the input and resulting in the learners overlooking the often-associated, redundant grammatical cues (N. Ellis, 2006).

One of the most influential and comprehensive models of L1 transfer is the Competition Model, first put forward by MacWhinney (1987) and MacWhinney and Bates (1989). The Competition Model posits that L1 sentence processing is governed by the detection of specific cues (e.g. word order, animacy, agreement), which hold differential weightings, or strengths, in individual languages depending on their availability, reliability, and validity in the input (MacWhinney, 1987, 2001, 2005; MacWhinney & Bates, 1989). In L2 acquisition, however, learners inevitably start out with cue weight settings which are close to those of their L1 (MacWhinney, 1987, 2005). Consequently it is argued that this can lead to learners overlooking or misinterpreting certain cues in the L2 input, due to the dominance of L1 cues, at least in the initial stages of L2 learning (MacWhinney, 2005).

From an usage-based perspective N. Ellis (2006, 2010) argues that L2 acquisition is mediated by selective attention. Selective attention results from the language learning apparatus having been tuned to the learner's L1. The system, therefore, becomes 'blind' to certain features in the L2 input, which can bring about L1 bias in subsequent estimations and statistical tallying (N. Ellis, 2006, 2010; N. Ellis & Sagarra, 2011).

From both the Competition Model and the usage-based perspectives, L1 transfer (of cues or selective attention) is thought to directly interfere in sentence processing in the L2, i.e. conversion of input to intake which becomes available for acquisition. However alternative theories have also been proposed, which, whilst agreeing that knowledge of the L1 does influence L2 acquisition, contest the manner in which the L1 operates (VanPatten, 2004a). In his Input Processing model, VanPatten (1996, 2004a) proposes that the L1 influences the developing L2 system by shaping the initial hypotheses which are generated with regard to the syntax of the new language. Intake, which is constrained by the principles of input processing (see section 2.3.3), is delivered to the developing system and is subsequently processed within the frame of the initial hypotheses generated based on knowledge of the L1 (VanPatten, 1996). Exposure to input which supports, or at least does not contradict, the L1 generated hypotheses would result in the developing system accepting them incorrectly. It is therefore argued that the L1 does not interfere during input processing, rather it acts as a "hypothesis generating knowledge source for the developing system" (VanPatten, 1996, p. 143).

Such a position would imply a role for instruction in optimizing both the input the learner is exposed to and their input processing in order to provide intake which is representative of the L2 and would result in the correct acceptance or rejection of the L1-based hypotheses. In contrast, based on the Competition Model, it can be argued that instruction can be utilised to maximise and optimise the input the learner is exposed to and push the learner to rely on the most reliable, valid and available cue in the L2 input, thereby bringing about a restructuring of the learner's cue hierarchy (MacWhinney, 2001; Stafford, Bowden, & Sanz, 2012). Similarly, N. Ellis (2010) proposes that form-focussed instruction can be beneficial in making all of the input 'count', rather than just the restricted sample which is often characteristic of biased L2 learning. By utilizing learners' explicit, conscious processing, the form-function mappings of novel L2 constructions can be consolidated (N. Ellis, 2005, 2010).

2.3.2.3 Consciousness-raising tasks

Given the observations regarding why some grammatical features may be more problematic for acquisition than others, and despite the controversy surrounding the Noticing Hypothesis and the question of whether awareness (at the level of noticing) is *required* for learning, it is important to consider whether increasing learners' awareness of problematic grammatical features could facilitate acquisition and if so, what kind of instruction would maximise learners' intake from the input (Reinders & Ellis, 2009).

Following the proposals put forth in Schmidt's Noticing Hypothesis, a considerable number of studies have investigated the issue of whether input-based instruction can be beneficial in helping learners to attend to problematic grammatical features in the input (e.g. Harley, 1998; S.-K. Lee & Huang, 2008; Loewen et al., 2009; Reinders & Ellis, 2009; Sharwood Smith, 1993; Simard, 2009; Spada, 1997; Stafford et al., 2012; Svalberg, 2012). Indeed Schmidt (2001) highlights that features such as infrequency, redundancy, and low saliency may make instruction a "practical (though not theoretical) necessity for successful language learning". Instruction can serve as a means of helping learning to organise the input they encounter, promote understanding and enhance natural acquisition processes (Schmidt, 1994, 1995).

A prominent form of input-based instruction utilises input enhancement tasks, which were originally referred to as "consciousness-raising" tasks (Sharwood Smith, 1991). Sharwood Smith (1991) defined consciousness-raising as "a deliberate focus on the formal properties of language with a view to facilitating the development of L2 knowledge" (p. 118). The term input enhancement was subsequently proposed as, crucially, it makes no assumption as to whether instruction results in changes to the learner's mental state (Reinders & Ellis, 2009; Sharwood Smith, 1993). Rather the focus of the input enhancement approach is on manipulation of the input in some way, with a view to increasing the learner's awareness of the target feature and subsequently maximising the potential for input to become intake (S.-K. Lee & Huang, 2008; Sharwood Smith, 1993; Simard, 2009). Further the aim of such tasks is for the learner to "arrive at an explicit understanding of some linguistic property or properties of the target language" (R. Ellis, 1997, p. 160); such a definition would imply the development of a certain level of conscious awareness of the target feature. In turn it is argued that this awareness can impact the learner's input perception and processing in order to potentially bring about acquisition (Sharwood Smith, 1993). Consequently there is debate as to the effectiveness of such activities for developing not only explicit knowledge of the language but also for bringing about an improvement in the learner's use of the target feature, as highlighted in the interface hypotheses (e.g. DeKeyser & Criado, 2012; N. Ellis, 2005; R. Ellis, 2005; Krashen, 1982).

Various different forms of input enhancement have been proposed. For example, within instruction utilising enriched input (or input flood) the learner is exposed to an increased number of exemplars of the target feature, with the aim that an increase in frequency will increase the saliency of the target feature (Reinders & Ellis, 2009). In comparison, methods such as enhanced input are designed to overtly draw the learner's attention to the target grammatical feature by emphasising it in some way, such as underlining, bolding, capitalization, or glossing (S.-K. Lee & Huang, 2008; Reinders & Ellis, 2009; Sharwood Smith, 1993; Simard, 2009). Sharwood Smith (1993) refers to techniques such as enriched and enhanced input as "positive" input enhancement, whereas methods such as corrective feedback would constitute "negative" input enhancement. Techniques can also be distinguished based on the degree of elaboration involved; approaches involving the use of metalinguistic terminology can be defined as highly elaborated, whereas those which do not call on metalinguistic knowledge would be unelaborated (Sharwood Smith, 1993). The following section will explore the efficacy of various forms of *positive* input enhancement (i.e. instruction concerned with manipulating the input to which the learner is exposed).

2.3.2.4 Enriched versus enhanced input

Overall, studies investigating different input enhancement techniques have presented mixed results. In general, findings have suggested greater effectiveness for more explicit measures and only small, if any, improvement resulting from more implicit methods such as input flood (Loewen et al., 2009; Marsden, 2006; Marsden & Chen, 2011; Reinders & Ellis, 2009; Spada, 1997; Svalberg, 2012; Trahey & White, 1993).

For instance, Trahey and White (1993) investigated the effects of enriched input for francophone learners of L2 English (target feature was adverb placement in English). Over the course of 10 days (total 10 hours) the learners participated in activities (e.g. stories, games), which provided naturalistic positive input containing an increased number of exemplars of the target feature. The learners received no instruction, or negative evidence (e.g. error correction) relating to adverb placement (Trahey & White, 1993). The results of four tasks (grammaticality judgement (GJT); sentence preference; sentence manipulation; oral production) indicated that the learners had learned that adverb placement between subject and verb is possible in English, however had *not* learned that placing an adverb between verb and object is UNgrammatical in English (Trahey & White, 1993). These findings suggested that positive evidence alone resulted in the learners developing only limited knowledge of the target feature (Spada, 1997; Trahey & White, 1993).

Similarly Loewen et al. (2009) found that intensive exposure to input containing the third person –s did not result in incidental acquisition of the target feature, as measured on tasks designed to tap into implicit (oral Elicited Imitation) and explicit (untimed GJT) knowledge. The learners completed a series of activities focused on the indefinite article *a*, however the materials were designed so as to also provide numerous exemplars of the target feature (Loewen et al., 2009, p. 272). It can be argued that the lack of improvement may have been due to the fact that during the intervention the learners were explicitly focussed on a separate grammatical form (indefinite article *a*), which may have distracted their attention from the target feature (third person –s) (Loewen et al., 2009). This conclusion would be in line with the proposal that engaging in 'dual-tasks' results in lower levels of learning (Schmidt, 1995). Further, factors such as learners' limited attention (N. Ellis, 2006; VanPatten, 2004a) and the redundancy of the target feature may have contributed to the learners' lack of improvement.

One explanation for the, generally, minimal improvement observed with eniched input is the fact that it is impossible to predict or guarantee that the learner will actually attend to the target feature (Svalberg, 2012). Despite the intention of enriched input being to increase the saliency of the target feature, increasing its frequency in the input may not be sufficient to push learners to notice the form. Jourdenais, Ota, Stauffer, Boyson, and Doughty (1995) found that enhanced input, in which the target feature (Spanish preterit and imperfect verb forms) was typographically highlighted, led to a higher incidence of learners' making reference to the target feature during a think aloud task as well as increasing their use of the form, compared to enriched input only. Consequently a number of studies (e.g. S.-K. Lee & Huang, 2008; Reinders & Ellis, 2009; Simard, 2009; Svalberg, 2012; J. White, 1998; L. White et al., 1991) have sought to investigate whether utilising input which has been enhanced (bolding, underlining, capitalization) leads to more substantial learning gains.

J. White (1998) investigated the effectiveness of three types of input on acquisition of the possessive determiners *his* and *her* by young francophone learners of L2 English: typographically enhanced input (bolding, italics, underlining) plus extended reading and listening; typographically enhanced input only; and unenhanced input flood. J. White (1998) hypothesised that enhanced input would lead to greater learning gains than unenhanced input as it overtly directs the learners' attention to the target feature. However no difference was found between the three groups. There was an increase in all groups' frequency of use of the target feature[13] following the intervention, but no improvement in the learners' accuracy when using the possessive determiners. These findings suggested that the typographical enhancement used did not result in the target feature being any more salient than for the unenhanced group (J. White, 1998).

Similarly Reinders and Ellis (2009) found that for learners' receiving enriched input, their use of the target feature (negative adverbs) increased to the same extent during the

[13] As measured on listening comprehension, multiple choice, oral picture description and passage correction tasks (J. White, 1998)

treatment as for the enhanced input group. The authors interpreted this finding as evidence that enriched input is equally as effective as enhanced input in increasing intake from the input. However, neither group made significant gains on a timed nor an untimed GJT, suggesting that instruction had only minimal impact on the learners' implicit and explicit knowledge respectively (Reinders & Ellis, 2009, p. 296).

S.-K. Lee and Huang (2008), in their meta-analysis of visual input enhancement studies, found a small overall effect size for enhanced over enriched input (d = 0.22). Further, analysis of the 20 studies included confirmed that there were discrepancies between the results of different studies. The stark methodological differences between various input enhancement studies (e.g. enhancement technique, grammatical feature and language targeted, learning measurements) could account for the observed disparity in results (S.-K. Lee & Huang, 2008; Reinders & Ellis, 2009; Simard, 2009).

In summary, investigations into the effectiveness of input-based instruction, centred on input enhancement, have revealed only small improvements in attention to and the learning of target grammatical features. Notably based on the findings of many of the studies it can be argued that employing a more explicit form of input enhancement (e.g. provision of grammatical rule or making noticing task essential) may be more beneficial than enriched or enhanced input alone. J. White (1998) proposed that the input enhancement utilised in their study may not have been explicit enough in order to induce noticing. Similarly Reinders and Ellis (2009) concluded that, due to the complexity of the target feature, the noticing instruction given to the enhanced input group may not have been sufficiently explicit in order to assist the learners to a greater extent than the enriched input alone. Finally, the minimal impact of input enhancement is perhaps not surprising given that such techniques do not require the learner to 'interact' in any with the enhanced input itself. Any noticing which may have occurred as a result of the input enhancement would not have pushed the learner to engage with the language and the target feature in any way (Svalberg, 2012). Consequently Svalberg (2012) argued that utilising an approach that incorporates tasks which encourage engagement with the target feature may enhance the effectiveness of providing enriched and/or enhanced input.

2.3.3 Input Processing theory

The Noticing Hypothesis (Schmidt, 1990, 1995, 2001) posits that learners need to notice (i.e. with awareness) a grammatical feature, in the context of meaningful input, in order for that feature to become intake and made available for acquisition. Alternatively Tomlin and Villa (1994) have argued that detection (without awareness) is the sufficient condition for acquisition. Although disagreeing as to the role of awareness, both of these theories propose that noticing, or detection, of elements of the *surface structure* of utterances, within meaningful input, is sufficient for acquisition to be made possible (Schmidt, 2001). Crucially, however, neither theory requires that any meaning (or function) is assigned to the grammatical form at the moment of noticing or detection (VanPatten, 2004a). In contrast, alternative theories have been put forth which argue that noticing needs to occur at a deeper level than that of surface structure alone.

VanPatten (1996, 2004a) proposes that in order for a learner to build a mental representation of the grammatical feature in question, the connection between its "referential real-world meaning (meaning) and how that meaning is encoded linguistically" (form) must be made (p. 10). VanPatten (2004a), hypothesizes that a learner's input processing (i.e. what they do to the input during comprehension) will determine which elements of the input are processed to become intake. Notably, here the term *processing* does not refer solely to the perception of a grammatical form in the input, but also to connecting that form with its meaning, or function, within the sentence during real time comprehension (VanPatten, 2004a). Successful processing, i.e. establishing the correct form-meaning connection (FMC), will result in the grammatical feature in question being converted to intake, which is subsequently available for further processing.

It should be noted at this point that, for VanPatten, intake is not equivalent to acquisition. Intake refers to the subset of input which is processed and subsequently held in working memory. In order for that intake data to be internalised into the learners' developing linguistic system, i.e. acquired, it needs to undergo further processing (VanPatten, 2002, 2004a). Importantly, however, conversion to intake does not guarantee acquisition (R. Ellis, 2001; VanPatten, 2004a); nevertheless further processing is not possible until the relevant elements of the input have first been converted to intake. VanPatten's input processing, then, is concerned with the first stage in the acquisition process, as denoted in **Figure 2.1** below:

$$\text{Input} \xrightarrow{1} \text{Intake} \xrightarrow{2} \text{Developing system} \xrightarrow{3} \text{Output}$$

1. input processing
2. accommodation, restructuring
3. access, monitoring, control

(Chen, 2009; J. F. Lee & Benati, 2007; VanPatten & Cadierno, 1993)

Figure 2.1 Schematic of SLA

Of note, however, is the fact that not all input becomes intake and in addition some input may be incorrectly processed (i.e. the wrong FMC is made) (VanPatten, 2004a). Therefore, in his theory of Input Processing, VanPatten (1996, 2002, 2004a) sets out a number of principles which detail how learners' processing of target language input is constrained. Accordingly these principles attempt to account for why certain grammatical features within the target language input fail to become intake and consequently are not acquired into the developing system.

2.3.3.1 Principles of Input Processing theory

VanPatten's Input Processing model consists of two core principles, which it is argued can account for why learners successfully process some grammatical forms, i.e. make the correct connection between the form and its meaning, but do not process others:

> ***Principle 1: The primacy of meaning principle***
> *Learners process input for meaning before they process it for form.*
>
> ***Principle 2: The first noun principle***
> *Learners tend to process the first noun or pronoun they encounter in a sentence as the subject or agent*
>
> (VanPatten, 2004a)

The present study is primarily related to Principle 2, therefore a detailed critique of Principle 2 and the accompanying claims made by VanPatten (1996, 2004a, 2007) is provided. A summary of subprinciples for principle 1 is provided in Appendix 1.

Before turning to Principle 2, it is important to consider the main concept underpinning VanPatten's model of Input Processing; namely that the inherent communicative value of particular elements in the input mediates whether or not attention is given to that element. Communicative value can be defined as "the meaning that a form contributes to overall sentence meaning" (VanPatten, 2002, p. 759) and constitutes the combination of a given form's inherent semantic value [+/– semantic value] and redundancy [+/– redundancy]. Crucially, semantic value is of primary importance over redundancy; with – semantic value forms containing little or no communicative value since they do not contribute to the overall meaning of the input (VanPatten, 2002).

Communicative value then is a key factor contributing to what learners pay attention to in the input. The higher a form's communicative value, the more likely it is that the form will be processed (i.e. FMC will be made) and made available as intake for acquisition (VanPatten, 2002, 2004a). Of note, given the context (foreign language classroom) in which the present study will be conducted is the fact that communicative value will also interact with frequency to affect whether or not a given grammatical form is attended to; those with low communicative value combined with infrequent occurrences in the input are likely to be 'doomed' to never being picked up by the learner (VanPatten, 2002).

2.3.3.2 Tenets of Principle 2

Principle 1 and its related subprinciples (Appendix 1) are often employed as explanations regarding the constraints on learners' processing of grammatical forms such as inflections or noncontent words. However syntax can also be conceived of as grammatical form, for example in terms of word order (WO) and how it conveys the relationship between the nouns and verb in a sentence (VanPatten, 2004a, 2007). English for instance is a strictly SVO language whereas Spanish and German allow more flexible word orders such as SVO, OVS, SOV, VOS[14]. Consequently comprehending the intended meaning of a speaker necessarily entails that learners assign both grammatical (e.g. subject versus object) and semantic roles (e.g. agent versus patient) within a sentence (VanPatten, 2004a).

In order to account for constraints on learners' processing of sentence structure, VanPatten (2004a) proposed Principle 2, the First Noun Principle (FNP), which claims that "learners tend to process the first noun or pronoun they encounter in a sentence as the subject or agent" (p. 15). Such a processing strategy results in the learner incorrectly assigning the subject/agent role to the first noun in the sentence, for example incorrectly interpreting sentence (1) as (2):

(1) *The cow was kicked by the horse*.
(2) *The cow kicked the horse*.

(VanPatten, 2004a, p. 15)

VanPatten (1996, 2004a) argues that the FNP constitutes the default strategy relied on by learners (from and of any language) when comprehending target language input. Accordingly a number of studies have been cited as evidence that learners have a strong tendency to rely on a WO (i.e. SVO) strategy, regardless of additional grammatical cues which may indicate otherwise.

In German for example, word order is flexible, and grammatical roles are most reliably marked by case marking. Consequently despite the word order having been reversed, both sentences (3) and (4) mean "the man (*Mann*) kisses (*küsst*) the woman (*Frau*)":

(3) SVO *Der Mann küsst die Frau.*
(4) OVS *Die Frau küsst der Mann.*

Previous research has demonstrated that learners often fail to attend to case marking cues, and therefore interpret sentence (4) as "the woman kisses the man". LoCoco (1987) for example found that when L1 English learners were presented with sentences such as (5) they tended to rely primarily on WO, interpreting it as "The truck pushes the car".:

[14] S = subject; V = verb, O = object

(5) OVS Den Lastwagen schiebt das Auto
The-ACC truck pushes the car

Across various conditions (simple transitive or ditransitive structures) LoCoco (1987) found that the learners incorrectly processed around 70% of OVS strings as SVO in response to aural stimuli. Similarly Jackson (2007) investigated intermediate L2 German learners' (L1 English) use of WO when interpreting sentences in a sentence level comprehension task. In line with LoCoco's observations, Jackson found that, despite having previously received instruction on case marking (prior to the study), the participants continued to misinterpret OVS strings in 50% of cases.

Evidence in support of the FNP has also been put forth in other languages. For example L1 English learners of L2 Spanish have been found to incorrectly interpret the Spanish object pronoun as a subject pronoun in OVS sentences (e.g. *'lo mató el león'* him killed the lion) (e.g. J. F. Lee, 1987; J.F. Lee & Malovrh, 2009; LoCoco, 1987; VanPatten, 1984).

Further studies by Allen (2000) and VanPatten and Wong (2004) demonstrated the difficulty learners can encounter when attempting to interpret causative (*faire*) structures in L2 French:

(6) *Jean fait promener le chien à Marie.*
(John makes to walk the dog to Mary.)
John makes Mary walk the dog.

(VanPatten & Wong, 2004, p. 98)

In both studies learners would tend to interpret the very first noun in the sentence (in this case *Jean*) as the subject of the verb *promener* (walk).

Such empirical findings have been interpreted as evidence in support of the existence of a universal processing strategy (FNP) which constrains learners' comprehension of target language input, particularly in the early and intermediate stages of language learning (VanPatten, 1996, 2004a). It can be argued that such a processing strategy can have a detrimental effect on learners' acquisition of languages which do not follow strict SVO word order (VanPatten, 2002). Erroneously relying on word order can lead to the learner overlooking and consequently not processing (i.e. not attaining the correct FMC for) key grammatical cues such as case-marking, passive constructions, and pronouns (VanPatten, 2004a). Indeed Jackson (2007) concludes that "although this preference for meaning-based or subject-first strategies often is an effective method for comprehending target language input, by not paying attention to particular grammatical forms in the input, learners may miss important cues that are necessary for interpreting a sentence" (p. 420). Further, rather than reaching the correct meaning via an alternative route (e.g. word order), very often the correct meaning is not reached at all (VanPatten, 2002).

2.3.3.3 Subprinciples of Principle 2

Whilst it is argued that the FNP is the default strategy with which learners will process target language sentences, there are a number of factors which may attenuate the FNP under certain circumstances (VanPatten, 2004a, 2007). Principle 2 can therefore be further sub-divided into three subprinciples.

> **Principle 2a: The lexical semantics principle**
> Learners may rely on lexical semantics, where possible, instead of word order to interpret sentences.

(VanPatten, 2004a, p. 16)

Lexical semantics in this sense refers to "the constraints on a situation imposed by the semantics of the verb involved" (VanPatten, 1996, p. 36), and which may circumvent learners' reliance on word order. For example animacy cues may assist (or equally in certain circumstances may hinder) the interpretation of sentences, such as in (10):

(10) *The fence was kicked by the horse.*

(VanPatten, 2004a, p. 16)

In the above example only one entity, *the horse*, is capable of performing the act of kicking, thereby eliminating, or reducing, the likelihood of *the fence* being misinterpreted as the subject / agent of the sentence.

In support of the role of lexical semantics in mediating interpretation of sentences in non-standard word order (i.e. first noun is not subject), LoCoco (1987) found that interpretation of German sentences considerably improved when the stimuli contained animacy cues (inanimate direct object). The percentage of erroneous interpretations of OVS sentences reduced from approximately 70% to 40%. Similarly Jackson (2007)'s participants relied on semantic information as well as word order when interpreting L2 German sentences, however did not utilise case marking (the most reliable cue to grammatical roles in German). Jackson (2007) found close to ceiling level performance on OS sentences containing only one animate noun from testing time 1, with the participants correctly interpreting 91.6% of the relevant sentence stimuli (compared to 50% of the OS sentences containing two animate nouns). In line with these findings Kempe and MacWhinney (1998) observed that their participants (L1 English) responded more quickly to (L2 German) sentence stimuli when the first noun was animate or the second noun inanimate, thereby aiding interpretation as the subject or object of the sentence respectively. Kempe and MacWhinney (1998) concluded that "learners of German consider semantic information immediately and regardless of whether an unambiguous case marker is present or not" (p. 567).

Additionally VanPatten (1996) cited evidence from a study by Issidorides and Hulstijn (1992) of English and Turkish[15] learners of L2 Dutch, as further support for Principle 2a. The participants were tested on their comprehension of sentence stimuli which were ordered adverb-verb-subject-object. Issidorides and Hulstijn (1992) found that when animacy cues were in conflict with the word order cues (i.e. the first noun was inanimate), the assignment of subject status to the first noun in the sentence dropped significantly for both groups of learners. Similar findings were observed in Gass's (1989, cited in VanPatten, 1996) study of English learners of L2 Italian, and Italian learners of L2 English; a reliance on lexical semantics was found to override word order in conflict sentences for both groups of learners.

A second factor, relating to the concept of semantics, can be identified as contributing to sentence interpretation, as summarised in principle 2b:

> **Principle 2b: The event probabilities principle.**
> Learners may rely on event probabilities, where possible, instead of word order to interpret sentences.
>
> (VanPatten, 2004a, p. 17)

In this context event probabilities can be defined as "the likelihood of one noun being the subject/agent as opposed to another" (VanPatten, 2004a). Whereas lexical semantics refers to whether a noun is able to perform the action denoted by the verb, event probability relates to whether it is *likely* for a given noun to perform the specified action in a real-world context. For example VanPatten (2004a) argued that learners are more likely to interpret a passive construction such as (11) correctly based on event probability, given that in the real world it is more likely for a dog to bite a man than vice versa:

(11) The man was bitten by the dog.

In support of this subprinciple, VanPatten (1996) cited a study by Bavin and Shopen (1989 (1989, cited in VanPatten, 1996), which demonstrated that lexical semantics and event probability can supersede a reliance on word order for active as well as passive sentences. In their study Bavin and Shopen observed that the participants (Walpiri speaking children) were less likely to rely on word order when event probability favoured the interpretation of one noun or the other as the agent of the sentence. A similar finding was observed for sentences in which the action was obligatorily performed by an animate noun, providing additional evidence as to the mediating role of lexical semantics.

The final subprinciple relates to the role of context in aiding interpretation of target language sentences. When learning within an immersion or naturalistic setting, it is unlikely that learners will be exposed to individual sentences in isolation; rather they will be embedded

[15] In Turkish the unmarked word order is SOV

within a wider communication act. Therefore the contextual information preceding a given sentence may act to attenuate the FNP and constrain possible interpretations of who did what to whom, within a given sentence:

> **Principle 2c: The contextual constraint principle**
> Learners may rely less on the First Noun Principle if the preceding context constrains the possible interpretation of a clause or sentence.

VanPatten and Houston (1998, cited in VanPatten, 2004a) observed that when learners interpreted sentences, in which the preceding context constrained possible interpretation of the target clause, learners reliance on the FNP was significantly diminished. For example in the sentence (12), interpretation of the underlined clause is constrained by the preceding information:

(12) Roberto está en el hospital porque <u>lo atacó María con un cuchillo</u>.
Robert is in the hospital because <u>him-OBJ attacked Mary with a knife</u>.

(VanPatten, 2004a, p. 17)

The information that Robert is in the hospital suggests that something has happened to him, thereby promoting the interpretation that Mary was the attacker. Similarly J.F. Lee and Malovrh (2009) observed that for beginner (L1 English) learners of Spanish, context was a significant factor which aided the learners' interpretation of OVS strings.

The three subprinciples outlined above account for the factors which may mediate learners' reliance on the FNP when processing target language input. An overreliance on cues such as lexical semantics, event probabilities, or contextual information, as well as on word order as predicted in the FNP, is likely to cause learners to overlook key grammatical features (e.g. pronouns, case marking) which may, more reliably, encode grammatical roles within a sentence (VanPatten, 2002).

2.3.3.4 Challenges to principle 2: Alternative explanations

Notably, it has been argued that rather than being a universal, default processing strategy, an over reliance on WO is a result of the learner transferring L1 grammatical cues to their L2 processing. As discussed in section 2.3.3.2, the Competition Model posits that learners utilise L1 processors, particularly in the initial stages of L2 learning, which results in their interpretation of target L2 input being constrained by the dominant cues (e.g. WO, case marking, animacy) in their L1. The findings of studies such as that of LoCoco (1987) and Kempe and MacWhinney (1998) can be taken as evidence in support of the Competition Model. The learners in both studies were L1 English speakers and accordingly were found to rely predominantly on WO when interpreting sentences in both German and Russian. These findings are in line with the predictions made in the Competition Model, since word order is the most reliable and valid cue to grammatical roles in English. In contrast it is argued that Italian learners rely most heavily on agreement cues, whereas German and Russian speakers utilise case marking cues, and in Spanish the prepositional object marker “a” provides the most reliable cue as to the object of the sentence (MacWhinney, 2001). Empirical evidence has been put forth, such as the study by MacWhinney, Bates, and Kliegl (1984) in which English, German, and Italian L1 speakers’ interpretation of English sentences, which contained conflicting cues, were investigated. In line with the predictions made in the Competition Model and the results of the studies above, the English speakers relied primarily on WO, whereas the German and Italian speakers made more use of animacy and agreement cues. Research has since been carried out demonstrating the differential strength of such cues for a range of languages (MacWhinney, 2001). In addition Isabelli (2008) investigated whether Italian learners’ of L2 Spanish had difficulty processing OVS structures (e.g. *Lo ve Maria*, *him sees Maria*), as would be predicted by the FNP. However, Isabelli found that the Italian learners were able to successfully interpret OVS structures in Spanish, therefore supporting an interpretation based on L1 transfer. Nevertheless it is important to note that Italian and Spanish share the object pronouns *lo* (him) and *la* (her), therefore VanPatten (2014) argued that lexical transfer may have attenuated any effects of the FNP in Isabelli’s study.

In response to an L1 cue-based explanation of the constraints of learners processing of L2 input, VanPatten (2004a) argues that the Competition Model and its associated empirical evidence do not necessarily constitute counterevidence to the existence of a default FNP processing strategy. Indeed VanPatten argues that within Competition Model research cues are deliberately put into conflict, for example in sentences in which word order and animacy assign grammatical roles to different nouns (e.g. *rock-throw-monkey*). Such cue conflict, in turn, elicits differential cue reliance for learners from different L1 backgrounds. However VanPatten (2004a) proposes that when faced with simple NVN constructions in which animacy and alternative cues remain neutral (e.g. *monkey-bite-baboon*), learners from a majority of L1s would select the first noun as the subject / agent. Consequently word order is put forward as the default, “core” processing strategy (VanPatten, 2004a, p. 24). Nevertheless it is also important to consider that for a majority of the languages (e.g. English, French, Spanish, Italian, German), which have thus far been investigated within the input processing framework, the dominant word orders tend to be SVO and SOV, although some may also allow object initial structures (e.g. OVS, OSV) (VanPatten & Wong, 2004). Consequently it is not possible to definitively conclude whether the findings of empirical studies, such as those cited above, are the result of a default, universal processing strategy, or transfer from the respective L1s.

2.3.3.5 Summary of principle 2

In summary, Principle 2 of VanPatten's (1996, 2004a) Input Processing model (FNP) posits that learners' parsing of L2 input is constrained by a reliance on word order. Namely, in the initial stages of language learning, learners tend to assign the subject role to the first noun encountered within a sentence. In the related subprinciples, VanPatten sets out a number of other factors which may circumvent this reliance on WO: lexical semantics (e.g. animacy), event probabilities, and contextual information. An important consequence of the FNP is that an overreliance on WO, as well as additional semantic and contextual cues can often result in the learner overlooking or misinterpreting key grammatical cues (e.g. case-marking, pronouns) in the input. Consequently learners will fail to process (i.e. make the correct FMC for) those grammatical features which are crucial in interpreting sentences which do not adhere to a strict SVO word order.

Notably however, alternative theoretical positions have also been put forth, such as the Competition Model, in which it is argued that the observed reliance on WO is due to the transfer of cues from the L1 (MacWhinney, 1987, 2001).

With regard to the present study, the participants are L1 speakers of English. Therefore, despite the ongoing theoretical disagreement, as to whether a default processing strategy or the L1 is responsible for learners' preponderance to label the first noun in a sentence as the subject, both Input Processing theory and the Competition Model would predict a bias towards a reliance on WO for these learners.

2.3.4 Processing Instruction

Processing Instruction (PI) is a type of input-based grammar instruction, which aims to aid learners in deriving richer intake from input by circumventing those processing strategies (detailed above) which normally constrain their processing of target language input (Wong, 2004a). Via the provision of structured input, PI seeks to push learners to attend to key grammatical forms, and their meaning or function within the input, thereby changing the way in which target language input is processed and subsequently altering the learners' developing system (VanPatten, 1996; VanPatten & Cadierno, 1993).

The structured input provided via PI can be defined as input which has "been manipulated so that learners are pushed away from less-than-optimal strategies" (Wong, 2004a, p. 37) and forms a fundamental cornerstone of the PI approach. It is also important to emphasise that PI does not include any production practice of the target grammatical form (VanPatten, 2002), since within the Input Processing framework, the production of output is not considered to contribute directly to the development of a learners' internal grammatical system (VanPatten, 2004a). Rather the focus of PI is on improving learners' processing and interpretation of target language input through attention to the relevant FMC (VanPatten, 2002). As such VanPatten (2002) characterises PI as a type of instruction akin to FonF or input enhancement (Sharwood Smith, 1993). Notably PI would not necessarily fit with Long's (1991) definition of FonF, given that PI is implemented with a predetermined linguistic focus and provides the learner with explicit information regarding the target feature. Nevertheless, contrary to the mechanical, drill-like activities which are characteristic of FonFS approaches, PI does have an overriding focus on meaning and is built on the premise of helping learners to attend to FMCs during the processing of meaningful, communicative input.

2.3.4.1 Core components of PI

The PI package contains four core components: (1) explicit information about the target grammatical feature; (2) information about the relevant processing problem; (3) referential activities; and (4) affective activities (VanPatten, 1996, 2002; Wong, 2004a). It is noted that the explicit information provided in element (1) is similar to the type of information provided in more traditional, output-based grammar instruction (although only one part of a paradigm is presented at any one time). Importantly however, VanPatten (1996) argues that components (2) to (4) are unique to the PI approach. The following subsections will offer a descriptive overview of each of the components. Previous research investigating the effectiveness of PI and its respective components will then be presented and critiqued in subsequent sections.

1) Explicit information about the target feature

Within the first component of the PI package, the learner is provided with information about the structural properties of the grammatical feature in question as well as relevant pedagogical grammar rules (VanPatten, 1996; Wong, 2004a). Crucially the meaning, or function, of the grammatical feature within target language sentences is highlighted in order to draw learners attention to the relevant FMC (VanPatten, 1996). It is also important to note that within PI only one grammatical form and its function should be the focus at any one time (Wong, 2004a). Within the explicit information provided, this target form is presented in juxtaposition to a comparison grammatical form, in order to highlight the FMC in focus. An example of the explicit information pertaining to the target grammatical feature (accusative definite article case marking in German) in VanPatten and Borst's (2012) study is presented below:

> In German, one way to tell what is the subject and what is the direct object is by looking at the definite article (the small word meaning "the") before the noun. This is especially true for masculine nouns. When a masculine noun is the subject, the definite article is ***<u>der</u>***. When it is the direct object, the article is ***<u>den</u>***.

2) *Explicit information about the processing problem*

The second component within the PI package constitutes the additional information provided about potentially problematic processing strategies (VanPatten, 1996), which, as mentioned previously, is unique to the PI package. Here learners are provided with information regarding the processing strategies which may result in incorrect processing (i.e. not establishing the correct FMC) of the target grammatical feature (Wong, 2004a). To illustrate how this component is operationalised in practice, the second half of the explicit information provided in VanPatten and Borst's (2012, p. 108) study is presented below:

> Word order in German is more flexible than in English. Whereas English is always subject-verb-object, German can be subject-verb-object and sometimes object-verb-subject. Compare the two examples below. Both mean "The woman sees the man."
>
> *Die Frau sieht den Mann.*
> *Den Mann sieht die Frau.*
>
> Thus, case markings on articles become important so that you do not misinterpret who does what to whom. Learners of German often rely on word order to determine who did what to whom, thinking the first noun is always the subject. But it may not be! If you see or hear ***den*** in front of a noun, that noun is not the "verb-er" and thus not the subject of the sentence.

In the above example, the processing problem identified is the First Noun Principle (section 2.3.3.2). Consequently information is provided to the learner, which is intended to shift the focus away from a reliance on word order and encourage a focus on the grammatical form when parsing target language input.

3) *Structured input: Referential activities*

It has been argued that the most important components of the PI package are the structured input activities (Sanz & Morgan-Short, 2004; VanPatten & Oikkenon, 1996) (section 2.3.4.4). These activities are specifically designed in order to give learners the opportunity to actively engage with structured input. The nature and purpose of this input can be described as follows:

> "input that is manipulated in particular ways so that learners become dependent on form and structure to get meaning and /or to privilege the form or structure in the input so that learners have a better chance of attending to it (i.e. learners are pulled away from their natural processing tendencies toward more optimal tendencies)"
>
> (VanPatten, 2002, p. 765)

Notably, the use of the term 'structured' highlights that in structured input activities, learners are not engaged in "free-flowing communicative discourse". Nevertheless such activities can be described as meaning-oriented since learners are encouraged to attend to the target grammatical form, and its meaning or function, during activities in which they see or hear language which expresses some meaning (VanPatten, 1996). Activities will typically include input provided via both written and aural modalities, i.e. through reading and listening tasks (VanPatten, 2002). Additionally, in the sense of structured input activities, 'manipulation' of the input involves not only providing extensive exposure to exemplars of the target grammatical form, but also removing those cues (e.g. lexical / content words, animacy, WO), which learners might otherwise rely on in order to interpret the target language sentences (J. F. Lee & Benati, 2007; VanPatten, 1996, 2002; Wong, 2004a). The focus is on pushing the learner to rely on a specific grammatical form to interpret meaning, thereby aiding learners in establishing the correct FMC (Wong, 2004a). Indeed, VanPatten (1996) states that "underlying all structured input activities is the push to get learners to make form-meaning mappings in order to create grammatically richer intake" (p. 55).

Referential activities require the learner to attend to the target grammatical form and its associated function in order to correctly interpret the sentence and reach the correct answer (in closed response format) (VanPatten, 2002; Wong, 2004a). Additionally referential activities have a right or wrong answer (VanPatten, 2002; Wong, 2004a) and typically learners will also be given brief feedback (correct / incorrect) either during or immediately following each referential activity (Sanz & Morgan-Short, 2004). Within referential activities, then, attention to the target FMC can be described as task essential, which is defined by Loschky and Bley-Vroman (1993) as a given grammatical form being required for successful completion of a task. An extract from a listening referential activity designed to teach masculine definite articles in German, is presented below (the learner was asked to select the picture that matched the aural sentence stimuli):

Der Junge küsst die Frau.

[Picture: boy kissing woman] [Picture: woman kissing boy]

Den Hund verfolgt die Katze.

[Picture: cat chasing dog] [Picture: dog chasing cat]

(Culman et al., 2009)

By presenting both SVO and OVS structures, the WO cue, which would often be relied on by learners to infer grammatical roles, was removed from the input. Consequently the learners were required to attend to the target grammatical form, the masculine definite article case marking, in order to correctly interpret the sentences and complete the activity (Culman et al., 2009). In this way referential activities can help learners bypass inefficient processing strategies and promote attention to a grammatical form and its FMC during input processing (VanPatten, 1996; Wong, 2004a).

4) Structured input: Affective activities

Affective activities are activities in which learners are required to "express an opinion, belief or some other affective response as they are engaged in processing information about the real world" (Wong, 2004a, p. 42). As such the inclusion of affective activities within the PI package was, in part, an endeavour to align PI with communicative language teaching, which often includes such affectively-oriented activities in order to foster learner-centred teaching and a focus on meaning (VanPatten, 1996; VanPatten & Cadierno, 1993). In affective activities, then, learners are required to respond in some way to sentences containing the target FMC, however, in contrast to referential activities, attention to the target FMC is not task essential. Rather the purpose is to reinforce the target FMC by providing extra exposure within meaningful input (Wong, 2004a). Consequently there is no right or wrong answer in affective activities (VanPatten, 1996, 2002; Wong, 2004a). Notably, however, as such affective activities do not guarantee that the learner will attend to the target FMC during completion (Marsden, 2006).

The two types of structured input activity (referential and affective) included in the PI package are claimed to be aligned with the predictions made in Input Processing theory regarding how input is converted to intake. An initial representation of the target FMC is established from the input (via referential activities), which is subsequently reinforced by exposure to multiple instances within meaningful input (via affective activities) (VanPatten, 1996, 2004a). Such an interpretation is in line with N. Ellis' (2002) Implicit Tallying Hypothesis, as well as Schmidt's (1990, 2001) Noticing Hypothesis, which propose that acquisition can occur as a result of the initial conscious registration of a grammatical feature and its connected meaning, coupled with extensive exposure to the relevant FMC within the input.

Guidelines for creating structured input activities

In order to achieve the purported improvement to learners' input processing, six guidelines are set out for the creation of structured input activities (VanPatten, 1996; Wong, 2004a).

(1) Present one thing at a time
(2) Keep meaning in focus
(3) Move from sentences to connected discourse
(4) Use both oral and written input
(5) Have learners do something with the input
(6) Keep the learners' processing strategies in mind

Guideline (1) is motivated by the proposal, which underpins Input Processing theory, that learners' attentional resources are limited. Consequently, only one FMC should be the focus of an activity at any one time, so as to "maximise intake efficiency" (Wong, 2004a). Additionally (6) underscores the unique characteristic of structured input activities; namely that they are motivated by a processing problem, which learners are known to encounter in relation to a particular grammatical feature. Consequently structured input activities should be created in order to tackle a particular processing difficulty. In order for an activity to be categorized as providing structured input it must adhere to the above guidelines.

2.3.4.2 Investigating the effectiveness of processing instruction

The original PI study by VanPatten and Cadierno (1993) set out to investigate its effectiveness in comparison to a more traditional form of grammar instruction. VanPatten and Cadierno (1993) defined traditional instruction (TI) as typically involving "explicit explanation of a form followed by controlled output practice … mechanical drills followed by meaningful and communicative drills" (Wong, 2004a, p. 45), and sought answers to the following research questions:

> 1. Does altering the way in which learners process input have an effect on their developmental systems?
> 2. If there is an effect, is it limited solely to processing more input or does instruction in IP also have an effect on output?
> 3. If there is an effect, is it the same effect that traditional instruction (TI) has (assuming an effect for the latter)?
>
> (VanPatten & Cadierno, 1993, p. 230)

The target grammatical feature in the study was object pronouns in Spanish and the processing problem predicted by the FNP (see section 2.3.3.2). The PI group in the study received explicit information about the target feature and related processing problem and completed structured input activities, such as the following:

> Listen as your instructor reads a sentence. Select the best interpretation from the English renderings.
>
> 1. a. My parents call me
> b. I call my parents
>
> Instructor reads aloud: *Me llaman los padres*
>
> (VanPatten & Cadierno, 1993, p. 231)

In comparison the TI group received explicit information about the target feature only, and then completed production drills ranging from mechanical to meaningful. A control group was also included who did not receive any instruction relating to the target feature. At pre- and post-test, all three completed sentence level interpretation and production tasks. The interpretation task required the learners to match the sentence to one of two pictures:

[picture: boy greeting girl]   [picture: girl greeting boy]

Al chico   *lo*   *saluda la chica.*
The boy-OBJ him-OBJpro greets the girl
"The girl greets the boy."

(VanPatten & Cadierno, 1993, p. 232)

The production task was a sentence completion task:

[picture: boy thinking about a girl]   [picture: boy phoning girl]

El chico piensa enla chica y entonces ___________________.
The boy is thinking about the girl and then ___*(he calls her)*___.

(VanPatten & Cadierno, 1993, p. 233)

VanPatten and Cadierno found that the PI group improved on both tasks, whereas the TI group improved only on the production task. It should be noted that the TI group did not complete any interpretation activities and nor did the PI group engage in producing the target feature at any point during the instruction (VanPatten & Cadierno, 1993). Based on these findings it was therefore argued that PI altered the way in which the learners processed the input, which subsequently affected the learners' developing system and therefore the knowledge accessible during production (VanPatten, 2002; VanPatten & Cadierno, 1993; Wong, 2004a). On the other hand, whilst improving the learners' production of the target feature, TI did not alter the way in which they processed the input and therefore did not improve their performance on the interpretation task (VanPatten & Cadierno, 1993), rather the TI group simply "learned to do a task" (VanPatten, 2002, p. 771). The findings of this seminal study have been taken as evidence in support of an Input Processing theory of SLA, i.e. that processing a novel item in the input, or processing input in a different way, facilitates acquisition more than production practice (VanPatten & Fernández, 2004).

Numerous replication studies have since been carried out with different languages and targeting a range of different grammatical features. For example studies have tested the effectiveness of PI for the English past tense '-ed' (Benati, 2005), French causative *faire* (VanPatten & Wong, 2004), Italian future tense (Benati, 2001), Spanish *ser* versus *estar* (Cheng, 2004), and Spanish past tense (Cadierno, 1995). The findings of such studies were in line with those of VanPatten and Cadierno; PI had a positive effect on both interpretation and production whereas TI only improved production ability. The authors of such studies have taken these findings as evidence of the generalizability of such results to other languages and grammatical forms (Benati, 2005).

Nevertheless it is important to note that a number of studies have found differing results to those presented above. For instance, in their self-labelled replication of VanPatten and Cadierno's study, DeKeyser and Sokalski (1996) found that their PI group outperformed their TI group on the comprehension task at post-test, but the opposite pattern was found for the production tasks. Similarly Allen (2000), in their study using the French causative, found similar gains for the PI and TI groups on the comprehension task, however on the production task the TI group outperformed the PI group. Such findings have been interpreted within the framework of Skill Acquisition theory (SAT) (DeKeyser, 2007; DeKeyser & Criado, 2012), which states, in line with general skill learning, that practice is required in order to develop individual language skills (e.g. comprehension, production) (DeKeyser & Sokalski, 1996). Further the knowledge, proceduralised and eventually automatized through practice is said to be "highly specific" (R. Ellis, 1999, p. 67); therefore contrary to Input Processing theory, SAT posits that input-based instruction will lead only to an improvement in learners' comprehension skills, whereas output-based instruction will develop production ability. In response to such claims, however VanPatten has argued that the observed differences in outcomes are due to the way in which PI has been operationalised in studies such as those by DeKeyser and Sokalski (1996) and Allen (2000). Specifically it has been argued that the instruction utilised in these studies did not meet the criteria for structured input activities, because, for example, a processing problem was not identified or attention to the relevant FMC was not made task-essential (TE) (VanPatten, 2002; Wong, 2004a).

Additional criticisms have arisen relating to the way in which TI has been operationalised in some PI studes. In VanPatten and Cadierno's study and subsequent replications TI has tended to consist of explicit information plus mechanical drills. Consequently it could be argued that the observed beneficial effects for PI may be due to the fact that it is fully meaning-based whereas the TI utilised was not (Farley, 2004b). Indeed different theories of SLA (DeKeyser, 2007; Krashen, 1982; MacWhinney, 2001; Schmidt, 1990; VanPatten, 1996), whilst disagreeing as to the role of grammar instruction, concur that some form of engagement (whether input- or output-based) with *meaningful* input is required for acquisition to occur. Nevertheless comparisons of PI and meaning-based output instruction (MOI)[16] have produced mixed results. Benati (2005), in a comparison of PI, MOI, and TI, with the English past simple tense, found equivalent gains for all three groups on a written production task, but that only the PI group significantly improved on the interpretation task.

[16] Meaning-based output instruction consists of explicit information coupled with communicative, rather than mechanical, output practice (Farley, 2004b)

Benati (2005), therefore, concluded that MOI is not beneficial in "bringing about similar effects to PI" (p. 84), namely altering learners' input processing. However, in their respective studies, Farley (2004b) and Morgan-Short and Bowden (2006) found that the PI and MOI groups improved an equivalent amount on the interpretation task, and in the case of Morgan-Short and Bowden (2006) the MOI group were actually found to outperform the PI group on the production task. Based on these findings it has been argued that the communicative nature of the MOI activities (e.g. asking learners to express opinions and beliefs using the correct form of the target feature) may have resulted in the learners producing extra incidental input for one another which was akin to that of the structured input activities in PI, thereby accounting for the observed equivalent benefits of MOI and PI (Farley, 2004b; Morgan-Short & Bowden, 2006). Critiques of PI studies have also highlighted methodological challenges, for example based on the nature of the tasks used to assess participants' learning. This issue will be addressed in more detail in section 2.3.5.

Whilst taking into account the disparity between the findings of various PI studies, it is important to consider that a majority of these studies have found a consistent beneficial effect for PI itself. It has consistently been shown to improve both learners' comprehension and production of key grammatical features in a number of languages. Notably the differences between previous studies have typically resided in how the comparison instruction (e.g. TI, MOI) has been operationalised and the resulting claims as to the superiority of PI. Indeed VanPatten (2004b) claimed that "although it is not clear that all output-based approaches always make a difference, PI always does" (p. 96). It is therefore difficult to deny the potential effectiveness of PI as a pedagogical tool for classroom-based foreign language teaching and learning, since at the very least it has been shown to be equally as effective as output-based instruction.

It is also important to consider that PI is an instructional technique claiming to operationalise Input Processing theory. As such, studies comparing PI to a form of output-based instruction are unable to provide conclusive support for an Input Processing theory of SLA, since the nature of the comparison instruction does not make it possible to falsify the theoretical claims within Input Processing theory (Marsden, 2006). In order to do so it is necessary to compare PI to alternative forms of *input-based* instruction, in order to find support, or counterevidence, to the fundamental claim of Input Processing theory and PI, that attention to a grammatical form *and* the meaning it encodes in the input is necessary in order for acquisition to occur.

2.3.4.3 Structured versus enhanced input

It can be argued that in some ways the structured input provided in PI is similar to the enriched or enhanced input provided though an input enhancement or consciousness-raising approach (Sharwood Smith, 1991, 1993), since structured input activities, and in particular affective activities, provide 'enriched' input containing numerous exemplars of the target grammatical feature. In addition the aim of both approaches is to make the target grammatical feature more salient to the learner.

VanPatten (1996), however, argues that despite such similarities PI is fundamentally distinct from input enhancement instruction. With input enhancement the aim is to draw the learners' attention to the target feature, for instance by increasing its frequency in the input or enhancing it in some way (S.-K. Lee & Huang, 2008; Sharwood Smith, 1993; Simard, 2009). PI, on the other hand, aims to provide repeated opportunities for consistently, correctly, connecting the target grammatical form with its function in the input and as such does so by structuring the input so that the learner is forced to attend to the FMC (J. F. Lee & Benati, 2007; VanPatten, 1996). Consequently VanPatten (1996) promotes an interpretation of PI as "not about raising learners' consciousness about grammatical form but instead as enriching their subconscious intake" (p. 85) (see section 2.3.4.6). In a similar vein VanPatten (1996) highlights that PI is not equitable to comprehension-based instruction; whilst PI does aim to improve learners' comprehension of target language input, it does so by first improving their processing of specific features within the input (i.e. by influencing what learners do with it).

A small number of studies have investigated whether combining PI with some form of visual (e.g. bolding, underlining, animating) or aural (e.g. spoken more loudly) enhancement increases the salience of the target feature and thereby the effectiveness of PI (e.g. Agiasophiti, 2013; J. F. Lee & Benati, 2007; Russell, 2012). Such studies have compared structured input activities with and without enhancement. J. F. Lee and Benati (2007), for example, found no differences between the two groups as measured on sentence-level interpretation and production tasks[17], concluding that the structured input activities included in PI, no matter how they are presented (i.e. with or without enhancement), are the main factor influencing learners' positive performance (p. 109). Similarly Agiasophiti (2013) and Russell (2012) found that, overall, their respective groups made comparable levels of improvement regardless of whether they completed activities +Enhancement or –Enhancement. Small localised effects of +Enhancement were observed though: Russell (2012) found that the +Enhancement group slightly outperformed the –Enhancement group on the production task at post-test, and a similar effect was observed for Agiasophiti's (2013) learners

[17] Interpretation task: indicating whether a given statement referred to an event in the present or the future. Production task: gap-fill (provide the correct conjugation of the verb)

on the interpretation task. Agiasophiti (2013) claimed that this result provided evidence that "the external typographical enhancement had a positive effect in making input salient internally and in getting further processed by the language learning mechanisms" (p. 172). However such a claim seems tenuous given that this difference was not observed consistently across all of the outcome measures and further had disappeared by delayed post-test. Additionally the overall equivalent improvement made by the +Enhancement and –Enhancement groups is perhaps unsurprising given the fact that both groups completed the same structured input activities. One of the aims of incorporating enhancement into structured input activities was to bring about correct processing sooner than without. Therefore it is suggested that the timing of the outcome measures included in the aforementioned studies may not have been suitable to pick up any such subtle differences between the two types of instruction. A more online measure may have been more successful in measuring whether the +Enhancement group had begun processing the target FMC sooner than the –Enhancement group.

Notably very few studies have sought to explore PI in comparison to an alternative form of input-based instruction; only two such studies have been found within the related literature. Firstly Marsden (2006) presented the findings of two experiments in which PI was compared to Enriched Input (EI). EI included the same explicit information as PI coupled with exposure to activities which consisted of an equal number of exemplars of the target feature[18], however attention to the target grammatical form and the relevant FMC were not task essential. The aim of this comparison was to test the fundamental claim underpinning PI, that in order for a grammatical feature to be acquired the connection between the target form and its meaning must be made (VanPatten, 1996, 2002, 2004a). The learners were tested on both listening and reading (comprehension) and writing and speaking (production) measures, which included discourse- as well as sentence-level production tasks. In experiment 1 Marsden's findings were consistent with previous PI studies; greater learning gains were observed for the PI group compared to the EI group on both the comprehension and production tasks. Similar findings were observed for the comprehension measures in experiment 2 (with another class and school); however on the production measures the EI group were found to make gains equal to the PI group. The findings of these experiments suggested that the structured input activities completed by the PI learners, and the required attention to the FMC, resulted in a significant improvement in their processing and interpretation, as well as production (experiment 1), of the target feature within the input. In contrast exposing the learners to enriched input did not result in the EI learners processing the verb inflections "in a way that

[18] L2 French verb inflections for tense, person, and number (Marsden, 2006)

aided learning, when numerous exemplars were presented to them in the input" (Marsden, 2006, p. 544). The results of the production tasks in experiment 2 were not necessarily contradictory of such a conclusion, rather Marsden (2006) proposed that the gains made by the EI group, as well as the PI learners in experiment 2, may have been due to differences in factors such the background teaching context, the nature of the measures used and the explicit information which was provided to both the PI and EI groups.

Similar findings to those of Marsden (2006) were observed in a study by Marsden and Chen (2011), which sought to investigate the differential benefits of the PI referential and affective activities respectively. PI affective activities and enriched input tasks can be considered as fundamentally similar, since both expose the learner to multiple instances of the target feature within meaningful input, however neither force attention to the relevant FMC when completing the activities (Marsden, 2006; Marsden & Chen, 2011). The learners (*N* = 120) in Marsden and Chen's (2011) study were divided into four groups: learners completing referential + affective activities (RA); referential activities only (R); affective activities only (A); and a control group. The target grammatical feature was the English past tense inflection –ed, and a timed GJT, gap-fill, and picture narration task were utilised in order to measure any learning gains between pre-, post- and delayed post-test. Both the RA and R groups were found to make significant gains on the GJT and gap-fill tasks at post-test, and these gains were sustained over the delayed post-test (6 weeks after the intervention), whereas the A and Control groups were not found to make any improvement between test times on any of the tasks (Marsden & Chen, 2011). In accordance with the findings of Marsden (2006), this study provided evidence that simply exposing learners to repeated instances of a grammatical form, without pushing them to notice or process it, did not result in learning (DeKeyser, 1995; Marsden, 2006; Schmidt, 1990; VanPatten, 1996, 2004a). Further, based on the observed, equivalent gains of the RA and R groups it can be argued that it was the task essential attention to the FMC provided within the referential (R) PI activities which were responsible for the observed effectiveness of PI as an instructional approach. The inclusion of affective activities for the RA group did not result in any additional gains.

It is important to note that enriched input, such as that utilised in Marsden (2006) and Marsden and Chen (2011), whilst providing opportunities for noticing to occur by increasing the frequency of the target feature, did not *require* the learner to notice or attend to the feature in any way in order to correctly complete the activities. It is therefore possible that the learners in the EI (with pre-practice explicit information) (Marsden, 2006) and A (without pre-practice explicit information) (Marsden & Chen, 2011) groups respectively may not have attended to the target grammatical feature at all (Marsden & Chen, 2011). Indeed Svalberg (2012) argues that even when some form of enhancement is utilised, the observed minimal effect on learning (e.g. S.-K. Lee & Huang, 2008) may be due to the fact that learners are not required to actively engage with the target form. Svalberg (2012), therefore, proposed that incorporating an element of engagement with language, for example by asking learners to deduce the relevant grammatical rule or to identify (e.g. circle, underline, highlight) the target feature, may enhance the minimal effects of providing enriched and / or enhanced input. In addition Marsden and Chen (2011) highlight that, whilst enriched input alone has not been found to produce equivalent learning gains to PI, the effectiveness of instruction which encourages the noticing of the target grammatical form only, i.e. without connecting it with the meaning or function it realises within the input, is not yet clear. Such a comparison would enable further insight into the strength of the claims put forth in Input Processing theory and operationalised through PI. Accordingly the present study aims to compare the effectiveness of PI referential activities with an alternative form of input-based instruction, which provides equivalent levels of exposure to the target feature, but which requires task-essential attention to the grammatical form only, without making the relevant FMC task-essential (RQ3).

2.3.4.4 Role of explicit information

As well as comparing PI to alternative forms of grammar instruction, a substantial body of research has sought to isolate whether the observed effectiveness of PI can be attributed to a specific component within this instructional approach i.e. the explicit information or the structured input activities. VanPatten and Oikkenon (1996) for example compared three groups of learners on their acquisition of object pronouns in Spanish: one group received full PI (explicit information plus structured input activities); one received explicit information only; and one completed structured input activities only (including correct / incorrect feedback). VanPatten and Oikkenon (1996) found that the full PI and structured input only groups substantially outperformed those learners who had received explicit information only, on both a comprehension and production task at post-test. Based on these findings it was concluded that the structured input activities, which forced the learners' attention to the target FMC within meaningful input, were accountable for the observed benefits (for comprehension and production) of PI, rather than the explicit information provided (VanPatten & Oikkenon, 1996). Numerous studies (e.g. Benati, 2004; Sanz, 2004; Sanz & Morgan-Short, 2004; Stafford et al., 2012) have since replicated VanPatten and Oikkenon's (1996) findings and provided further evidence to suggest that the provision or absence of explicit information (prior to instruction, or during instruction via explicit feedback e.g. Sanz & Morgan-Short, 2004) about the target feature does not mediate the effectiveness of the structured input activities in improving learners' processing of target language input.

Nevertheless opinion remains divided as to whether the provision of explicit information can offer additional benefits for learning. Norris and Ortega (2000), in their meta-analysis of effect of instruction research, concluded that explicit instruction (e.g. activities which included explicit information or resulted in learners inducing a grammatical rule) tended to result in greater learning gains (as measured on controlled tests) than more implicit measures. Similarly based on the findings of their respective studies Reinders and Ellis (2009) and White (1998) proposed that the impact of input enhancement could have been improved if it had utilised more explicit techniques (i.e. explicit information).

Further, contrary to the findings of VanPatten and Oikkenon (1996) and other replication studies, Farley (2004a) found that full PI resulted in greater gains in comprehension and production of the Spanish subjunctive than structured input activities alone. Farley (2004a) contributed this discrepancy in findings to the nature of the target grammatical feature, which is arguably more “opaque” and semantically non-transparent than that of VanPatten and Oikkenon’s study (Spanish object pronoun). When completing structured input activities, without the provision of explicit information, learners would be required to conduct an “item-by-item analysis” in order to induce the correct FMC. Explicit information could, therefore, serve to help learners “see the connections in the structured input analysis more quickly” (Farley, 2004a, p. 238), which may be particularly beneficial for more complex grammatical items. In line with Farley’s findings, in more recent years a number of studies utilising the online measure ‘trials to criterion’ have demonstrated that the inclusion of explicit information within PI can function as a means of speeding up learning and result in learners correctly processing the target FMC faster than without (e.g. Culman et al., 2009; VanPatten & Borst, 2012). Consequently it has been argued that “distilled and focussed” explicit information which is “portable” enough to be accessed during processing can aid in the internalization of aspects of the grammar (Culman et al., 2009).

As demonstrated above studies have produced mixed results as to the role of providing explicit information in PI and instruction more generally. Nevertheless it is important to note that structured input activities alone have consistently been shown to lead to gains in learner’s comprehension and production of numerous grammatical features. It has therefore been argued that the task-essential attention to the FMC which is enforced via structured input activities is the “*necessary and perhaps sufficient* component of PI that leads to form-meaning connections in instructed SLA” (Farley, 2004a, p. 238; see also Marsden & Chen, 2011; Sanz & Morgan-Short, 2004). Notably DeKeyser et al (2002) argue that providing learners with structured input activities alone may be effective because it still results in explicit rule learning; the learners are continuously provided with correct/incorrect feedback which will lead to the learners inducing the FMC. Structured input activities alone, then, offer a type of explicit inductive approach, whereas full PI (i.e. including explicit information) could be considered an explicit deductive approach. Building on this debate, section 2.3.4.6 considers in more detail the nature of knowledge developed following PI.

2.3.4.5 Processing instruction for young learners

As discussed in section 2.1, age is an important contributory factor to the success of language acquisition. Further previous research has demonstrated that explicit instruction can also be useful for children learning a second or foreign language within the classroom. Notably, however, a majority of PI research has tended to target older learners and adults, and relatively few studies have explored the effects of PI for younger, primary school-aged learners (i.e. younger than 11), who could also be identified as "pre-critical-period" in line with the CPH. Nevertheless, based on the overwhelmingly positive effect of PI observed with adult learners, Benati and Lee (2008) put forward the Age Hypothesis which states that:

> "PI will be just as effective as an intervention with younger learners as it is with older learners." (p. 168)

A small number of studies have investigated the effectiveness of PI for younger learners (e.g. Angelovska & Benati, 2013; Laval, 2013; Marsden & Chen, 2011; Mavrantoni & Benati, 2013) and have provided evidence, which supports the claim set out in the above hypothesis. Laval (2013) for example found that PI was successful in improving young learners' (aged 9 to 10 years old) comprehension and production of the French imperfect. Similarly Mavrantoni and Benati's (2013) study revealed that PI improved their participants' (aged 8 to 10 years old) performance on interpretation and production tasks for the English third person singular inflection '-s'. In their study Angelovska and Benati (2013) also found greater learning gains following PI than TI for young learners (mean age 10.5 years) in their interpretation of the English past tense '-ed'. These findings were in line with those of Marsden and Chen (2011), whose participants were aged 12 years old.

These initial investigations into the effectiveness of PI for young learners have produced promising results. However it is important to note that the sample sizes of two of these studies were relatively small; Laval (2013) had only 14 participants, and Mavrantoni and Benati (2013) had 20 participants in the young learner group. In addition, with the exception of Marsden and Chen's study, only sentence level interpretation and / or production tasks were employed to measure learning in the above studies. Therefore further research is needed in order to provide more robust evidence of the effectiveness of PI for younger learners by recruiting larger sample sizes, employing a more comprehensive battery of outcome measures and exploring the generalizability of these findings for a wider range of languages and grammatical features.

2.3.4.6 Does PI develop implicit (-awareness) as well as explicit (+awareness) knowledge?

In his explanation of Input Processing theory, which is argued to constitute the theoretical basis of PI, VanPatten quite openly sidesteps any discussion as to the role of consciousness and awareness within this framework. VanPatten (1996) argues that "although my position is that awareness is probably a part of input processing at least initially, it is not necessary for positing the strategies described" (p. 46) (i.e. the principles described in section 2.3.3). Further Input Processing theory posits that the conversion of input to intake occurs once a given grammatical form has been "detected" within the input and the correct FMC has been made (VanPatten, 1996). This detection, as defined by Tomlin and Villa (1994), can be disassociated from awareness. Likewise VanPatten and Fernandez (2004) proposed that the correct FMC is made when the learner's internal processors notice the mismatch between the intended meaning and the meaning processed and argued that "processors … by definition perform their computations without awareness" (p. 280). Further VanPatten (2002) claimed that "the data contained in intake automatically make their way into the developing mental representation" (p. 762); this reference to automaticity would again suggest that these processes are occurring without conscious attention or awareness (see also VanPatten, 2007; VanPatten & Fernández, 2004). Such a position would suggest that input processing relates to a more implicit form of learning which happens without awareness and indeed VanPatten (2002) has argued that SLA results in the development of an implicit system.

As the pedagogical operationalization of Input Processing theory, the findings of PI studies have tended to be interpreted within the Input Processing framework. Benati (2001) stated that "the ultimate scope of processing instruction is not about raising conscious awareness about a grammatical form but to make the learner appreciate the communicative function of a particular form and consequently enrich the learner's intake" (p. 99). However, as highlighted by DeKeyser et al (2002), it is not made clear how this "appreciation" can occur without consciousness.

In addition, the learning gains made following PI have often been interpreted as evidence that the instruction served to optimize learners' processing of target language input and thereby caused changes to the underlying developing system (e.g. Benati, 2001, 2004; Benati, 2005; J. F. Lee & Benati, 2007; VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996; VanPatten & Wong, 2004). In contrast it has been argued that more traditional output based instruction, which utilizes explicit rule production practice, leads to the development of a "different kind of knowledge system" (VanPatten & Cadierno, 1993, p. 238). As such it can therefore be inferred that the learning gains made following PI, interpreted within the Input

Processing framework, have been taken by VanPatten and colleagues as evidence of the development of a more implicit form of knowledge (as noted in Marsden & Chen, 2011).

Nevertheless it is important to note that a key component within the PI package provides learners' with *explicit* information about the target feature, which it is argued can speed up learners' processing of key FMCs (Culman et al., 2009; VanPatten & Borst, 2012). Therefore the observed benefits of PI could be interpreted as support for the weak interface hypothesis and the facilitative role of explicit knowledge in the development of implicit knowledge (see section 2.2.5). DeKeyser et al (2002), however, argue that even without the provision of explicit information, through the completion of structured input activities learners induce the target grammatical rule, which in turn results in explicit learning. Indeed despite their instructional treatment providing no explicit knowledge, Marsden and Chen (2011) observed (via Principal Component Analysis) that the knowledge gained by their learners following referential activities tended to reflect explicit knowledge. In addition PI has been characterised as providing practice with explicit rules rather than acquired implicit knowledge (De Jong, 2005). Consequently DeKeyser et al (2002) proposed that "very little if any research on PI can even claim to address acquisition and not just the learning of monitored knowledge" (p. 819).

There continues to be extensive debate as to whether PI results in explicit learning and therefore the development of explicit knowledge or whether it does in fact cause changes in the learners' underlying linguistic system. Crucially, however, claims as to the nature of the knowledge promoted via PI, as well as other instructional approaches, are constrained by the way in which learning is measured within a given study; an issue to which we now turn.

2.3.5 Methodological issues in effect of instruction research

An important methodological issue which concerns PI studies, as well as effect of instruction research more generally, relates to the nature of the instruments used to measure learning and subsequent claims which are made as to the nature of the knowledge influenced through instruction. A majority of PI studies (Benati, 2001, 2004, 2005; Cadierno, 1995; Farley, 2004a, 2004b; VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996; VanPatten & Wong, 2004) have tended to rely on sentence-level interpretation and production tasks. Similarly in their meta-analysis Norris and Ortega (2000) found that a majority of the grammar instruction studies analysed (90%) employed measures such as *metalinguistic judgement tasks* or *constrained constructed response* (e.g. gap-fill, sentence transformation), which only required the learner to produce short segments of the target language. Such measures are likely to promote the use of explicit rather than implicit knowledge, since they require production of the target feature within a highly controlled linguistic context (Norris & Ortega, 2000, p. 486).

It has therefore been suggested that “the case for explicit instruction has been overstated” (Doughty, 2003, p. 274; see also Truscott, 2004). Further such measures are arguably not valid for testing whether the underlying, developing system has been changed in any way (De Jong, 2005; Doughty, 2004; R. Ellis, 2009c), a claim which is central to interpretations of the observed learning gains resulting from PI. Rather, the sentence-level tasks used in many PI studies simply test a learners’ metalinguistic, declarative knowledge about the language (Doughty, 2004). Consequently an increasing number of studies have promoted the use of discourse-level, as well as timed, tasks in order to provide more valid measures of spontaneous language use, which are thought to be more representative of the learner’s linguistic competence, i.e. implicit knowledge (R. Ellis, 2005, 2009b; Marsden & Chen, 2011).

Whilst observing a clear effect of PI for sentence-level comprehension and production tasks (e.g. Benati, 2001, 2004; Benati, 2005; Cadierno, 1995; Farley, 2004a, 2004b; VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996; VanPatten & Wong, 2004), more mixed findings have been observed for discourse-level measures (e.g. Cheng, 2004; Marsden, 2006; Marsden & Chen, 2011; Sanz & Morgan-Short, 2004). Sanz and Morgan-Short (2004) for example found that the learners’ use of the Spanish object pronoun significantly improved in a written video retelling task following instruction via PI. In contrast Marsden and Chen (2011) observed no significant change over time in their participants’ use of the past tense English –ed, when assessed using an oral picture narration task and a structured conversation. Further, two experiments by Marsden (2006) utilised similar tasks to Marsden and Chen (2011), however found mixed results. The findings of experiment 1 (class 1) revealed that the learners’ performance was approaching significance on the tasks, whereas no effect was found in experiment 2 (class 2). This discrepancy in findings could be accounted for by the modality of the respective tasks. Sanz and Morgan-Short utilised a *written* discourse task, which arguably could have served as a measure of more explicit rather than implicit knowledge since it was untimed and the learners would have had the opportunity to reflect on their language use whilst completing the activity (R. Ellis, 2005, 2009b). In contrast, the oral tasks utilised by Marsden (2006) and Marsden and Chen (2011) would not have afforded the same opportunity for reflection.

In terms of effect of instruction research more generally, Norris and Ortega (2000) have found similarly mixed results for the small number of studies (16%) employing “free constructed response” tasks (p. 470). However, in their meta-analysis, Spada and Tomita (2010) observed the largest effect size for “free” measures (involving spontaneous, unanalysed use of the target feature) following explicit instruction on complex grammatical features (p. 285). Likewise, R. Ellis (2002) reviewed 11 studies which examined the effect of form-focussed instruction on free production and found that instruction did appear to impact on the development of learners' implicit knowledge. However, contrary to Spada and Tomita (2010), R. Ellis (2002) found that instruction aimed at simple structures was most effective, although he proposed that instruction on more complex features can be effective provided that the feature is readily available in the non-instructional input.

Nevertheless it is important to acknowledge that questions still remain regarding whether tasks classified as "free" are in fact true measures of spontaneous unanalysed language use, and therefore more implicit knowledge (Spada & Tomita, 2010). Further research is therefore needed in order to gain deeper insight into the effect of instruction on different knowledge types.

2.4 Grammatical sensitivity and language learning

2.4.1 Defining grammatical sensitivity

When investigating an instructional technique, it is important to consider the role that individual differences (e.g. age of acquisition, aptitude, motivation, working memory) can play in mediating the effectiveness of instruction (Dörnyei & Skehan, 2003; Larsen-Freeman, 2006, 2014). Aptitude, "the specific talent for foreign languages which exhibits considerable variation between learners" (Dörnyei & Skehan, 2003, p. 590), is thought to be one of the most consistent predictors of success in language learning (Carroll, 1971; Dörnyei & Skehan, 2003). Carroll and Sapon (1959) identified four sub-components within the aptitude construct; phonemic coding ability, grammatical sensitivity, inductive language learning ability, and associative memory. Of relevance to the present study is the sub-component *grammatical sensitivity*, which has been defined as learners' ability "to recognize the grammatical functions of words in sentences" (Carroll, 1981, p. 105) and "to detect relationships among words" (VanPatten & Borst, 2012, p. 96). Further Skehan (1998) proposed a more general sub-component of *language analytic ability* which is defined as "the capacity to infer rules of language and make linguistic generalizations and extrapolations" (p. 204) and comprises both grammatical sensitivity and inductive language learning ability.

Although grammatical sensitivity does not require metalinguistic knowledge, i.e. knowledge of metalinguistic terminology, it presupposes a conscious meta-awareness of grammatical constructs (Krashen, 1981). Accordingly a number of studies have observed a significant relationship between grammatical sensitivity and learners' metalinguistic awareness, for example in terms of their ability to identify and or describe L2 errors (e.g. Alderson, Clapham, & Steel, 1997; Ranta, 2002; Roehr, 2006). Consequently it has been argued that grammatical sensitivity (along with inductive learning ability) relates primarily to conscious language learning (e.g. learning which takes place in instructed settings), rather than acquisition (Krashen, 1981). R. Ellis (2004) proposed that grammatical sensitivity can be viewed as an "essential ability underlying the development of explicit knowledge" (p. 251). Further Robinson (1997) argued that the Words in Sentences test, the grammatical sensitivity sub-test of the Modern Language Aptitude Test (MLAT) (see section 3.6.7), is a measure of ability to "control access to the learned, but not the acquired, system" (p. 54).

Skehan (1998, 2002) has proposed that the four components of foreign language aptitude can broadly be related to stages of information processing in SLA, with grammatical sensitivity relating to the stages of pattern identification and pattern restructuring and manipulation[19]. As such, aptitude, and its respective sub-components, may influence learners' processing of target language input by mediating what learners attend to (Dörnyei & Skehan, 2003; Skehan, 2002). In the case of grammatical sensitivity this may relate to whether a learner is able to correctly attend to the function of certain grammatical elements within the input. Consequently Robinson (2002) proposed that "input processing instruction, as described by VanPatten (1996) may be a technique for inducing focus on form that is differentially affected by the fourth aptitude complex […] particularly the grammatical sensitivity component of what I have termed metalinguistic rule rehearsal" (p. 131).

2.4.2 Grammatical sensitivity and instruction

A number of studies have observed a relationship between grammatical sensitivity and achievement following instruction (e.g. de Graaf, 1997; Robinson, 1997; VanPatten & Borst, 2012; VanPatten, Borst, Collopy, & Qualin, 2013). Robinson (1997), for example, compared the effect of grammatical sensitivity on learning[20] under four conditions varied by explicitness: instructed (explicit information given); rule-search (instructed to search for rules within sentence stimuli); implicit (exposure to stimuli without attention to the relevant rules and questions relating to the position of certain words); incidental (exposure to the same input plus a meaning-related task). Robinson (1997) found that grammatical sensitivity correlated significantly with learners' GJT performance for all conditions, except the incidental condition. Additionally the largest correlations with grammatical sensitivity were observed for the implicit condition. Further, awareness at the level of Looking for Rules and Ability to Verbalise rules were found to be predictors of superior learning for the Implicit group. Robinson (1997) therefore argued that "conscious awareness facilitated successful learning in this condition" (p. 82), supporting the proposal that the construct of grammatical sensitivity is related to explicit learning (DeKeyser, 2000; Krashen, 1981; Robinson, 1997). Further, it has been argued that aptitude relates to learning under conditions in which there is an overarching focus on form,

[19] See Dörnyei and Skehan (2003, p. 597) for a detailed list of the respective SLA stages and corresponding aptitude constructs

[20] Simple rule: subject-verb inversion with fronted-adverbial ("Into the house John ran/ran John")
Complex rule: pseudo-clefts ("Where Mary and John live is in Chicago not in New York") (p.59)

rather than a focus on meaning, such as the incidental condition in Robinson's study (de Graaf, 1997; Dörnyei & Skehan, 2003; Ranta, 2002).

In line with this proposal, VanPatten and Borst (2012) found a significant, albeit weak, correlation between grammatical sensitivity and performance on (PI) referential activities following explicit information (+EI), but not for the learning condition in which no explicit information was provided (-EI). VanPatten and Borst suggested that the -EI condition, in which the learners completed referential activities only, may have served as more meaning-focussed instruction, in line with Robinson's (1997) incidental condition. In addition given that the +EI group were found to start correctly comprehending the target FMC[21] sooner than the –EI group, VanPatten and Borst (2012) proposed that grammatical sensitivity may be one factor effecting a learners' ability to utilise explicit information during a processing task.

2.4.3 Grammatical sensitivity and young learners

In a study investigating the learning of L2 English in a naturalistic setting[22], DeKeyser (2000) observed a relationship between verbal analytical ability[23] (i.e. grammatical sensitivity) and the achievement of near-native speaker competence (as measured on a GJT) for those learners who were adult immigrants but no relationship for those learners who were childhood immigrants. DeKeyser (2000) argued that those adult learners with high verbal aptitude were able to utilise "explicit learning mechanisms to bypass the increasingly inefficient implicit mechanisms" (p. 518). Similarly, in two studies carried out with adolescent learners learning within an immersion classroom (Harley & Hart, 1997) and on an intensive bilingual exchange programme (Harley & Hart, 2002), Harley and Hart observed a relationship between grammatical sensitivity, and the L2 (French) proficiency of late immersion learners (intensive L2 exposure began in grade 7; age 12 to 13), however no relationship was observed for those learners who had begun instruction in early childhood (from grade 1; age 6 to 7). Harley and Hart (1997) concluded that "when intensive L2 exposure begins around adolescence, language learning will tend to depend on different cognitive abilities from those that early learners rely on, with analytical language ability being more intimately involved in L2 success for later learners" (p. 395).

The implication of such findings, then, is that aptitude plays less of a role in child language learning since young learners are able to rely on more implicit language learning mechanisms (DeKeyser, 2000, 2003). Notably, however, the learning context of the above studies constitutes a key factor in this suggestion. As highlighted in section 2.1.2.2, within the

[21] Nominative / accusative case marking on definite articles in L2 German (VanPatten & Borst, 2012)
[22] Participants were Hungarian immigrants to the USA (DeKeyser, 2000)
[23] Measured using the Words in Sentences subtest (a measure of grammatical sensitivity) from Hungarian Language Aptitude test (adaptation of MLAT) (DeKeyser, 2000, p. 509)

instructed, foreign language classroom setting younger learners, as well as adolescent and adult learners, may be more reliant on more explicit, problem-solving processes due to limited exposure to input and the use of explicit, form-focussed instructional techniques. In addition as highlighted by Philp et al. (2008), middle and older childhood (from age 7+) are characterised by the development of greater metalinguistic awareness and "a greater capacity for abstract thought, including language analysis" (p. 6). Consequently it is important to consider whether grammatical sensitivity, may also have a bearing on the learning outcomes of younger learners learning within the foreign language classroom. Indeed in their study of the teaching of Esperanto for 8-9 year old L1 English children, Tellier and Roehr-Brackin (2013a) found that language analytical ability was a consistent predictor of L2 achievement[24]. The present study will therefore incorporate a measure of grammatical sensitivity in order to explore the relationship with the learners' performance following instruction.

2.5 Target grammatical feature

The target grammatical feature in the present study is accusative case-marking on masculine definite articles in German. In German the correct choice of definite article for a given noun is determined by both the gender and the case of the noun:

Table 2.1: Definite article case-marking in German

	Masculine	Feminine	Neuter
Nominative (subject)	**der**	die	das
Accusative (object)	**den**	die	das
Dative (indirect object)	dem	der	dem
Genitive (possessive)	des	der	des

(Culman et al., 2009)

In German grammatical role is encoded in the case-marking on the definite, or indefinite, article. Notably in the accusative case, this proves ambiguous for feminine and neuter articles, which are the same as their nominative versions, <u>*die*</u> and <u>*das*</u> respectively (VanPatten & Borst, 2012). For masculine nouns however there is a clear distinction between the nominative and accusative cases. In addition, although the standard word order utilised in German is SVO, alternative word orders (e.g. OVS) are possible. The subject, object, and indirect object can move freely within the sentence provided that the verb remains in position two (Culman et al., 2009; Jackson, 2007). As noted in section 2.3.3.2, the FNP would predict that when faced with a sentence in which the word order has been reversed (OVS), learners would incorrectly assign

[24] As measured on tests of reading, writing, and listening relating to core vocabulary and structures covered in the treatment (Tellier & Roehr-Brackin, 2013a, p. 12)

the subject/agent role to the first noun in the sentence. Notably alternative theories such as the Competition Model (MacWhinney, 2001; MacWhinney & Bates, 1989; MacWhinney et al., 1984) would also predict that learners may encounter difficulty in interpreting sentences carrying non-standard word order if their L1 relies heavily on word order cues for the assignment of grammatical roles. L1 German speakers have been shown to rely on case-marking cues wherever possible when interpreting German sentences, as it is the most reliable cue to the correct assignment of grammatical roles, since, unlike in English, word order is flexible in German (Jackson, 2007; MacWhinney & Bates, 1989; MacWhinney et al., 1984). Speakers of L1 English on the other hand rely predominantly on word order and have been shown to have difficulty fully mastering the German case-marking system, possibly due to an overreliance on their L1 processing strategy (Jackson, 2007). In addition learners have tended to dismiss the importance of learning to correctly use (interpret and produce) case-markings since "L1 German speakers will understand what they mean, even if their case markings are not correct when they speak or write" (Jackson, 2007, p. 419). However an overreliance on word order (whether due to a default processing strategy or as an effect of the L1) constitutes an important problem for English learners of L2 German, even for those who are considered to be at an advanced level (Culman et al., 2009; Jackson, 2007; LoCoco, 1987; VanPatten & Borst, 2012).

Case-marking has often been perceived as a complex and abstract set of rules with little or no communicative value (Jackson, 2007). However a small number of studies (e.g. Culman et al., 2009; VanPatten & Borst, 2012) have demonstrated that it is possible to improve learners' comprehension and use of case-marking (specifically accusative case-marking on masculine nouns) through increasing their awareness of the feature and the importance of correctly interpreting case-marking cues and the relevant FMC. The present study therefore seeks to contribute to research in this area.

Traditionally the marking of a particular case, for all three of the genders is dealt with together in the language classroom (Culman et al., 2009; Edexcel GCSE German textbook, Lanzer & Wardle, 2009; Zoom Deutsch 1 textbook, Schicker et al., 2011). However, in PI, drawing on a limited attention model, only one form and its function should be the focus at any one time, in juxtaposition to a comparison FMC (J. F. Lee & Benati, 2007; VanPatten, 1996; Wong, 2004a). Therefore, in the present study, the focus will be on teaching the accusative case-marking, in comparison to the nominative case-marking, of masculine articles only.

It should also be noted that the current primary school curriculum for England (DfE, 2013c) and related schemes of work (e.g. QCA, 2007b) include the teaching of basic grammatical concepts and forms in the target language. For example the QCA scheme of work for KS2 German (aged 7 to 11) introduces definite articles from Unit 11 and accusative indefinite articles in Unit 8. Therefore the target grammatical feature chosen for the present study constitutes a key aspect of German grammar, which is included in non-statutory guidance for primary school level foreign language teaching in the UK.

2.6 Rationale and research questions

The findings of this study will contribute to a number of key debates regarding explicit grammar instruction and its role in SLA. Firstly, as yet, a relatively small amount of research investigating the role of explicit knowledge and the effect of explicit grammar instruction has been carried out with young learners. Indeed in their meta-analysis Norris and Ortega (2000) found that 79% of the studies analysed were conducted with adult learners, and only one study with elementary, or primary school-aged learners. Similarly only six of the 41 studies included in Spada and Tomita's (2010) meta-analysis involved elementary school-aged learners. One reason for this bias towards older learners in effect of instruction research is that younger learners are thought to be able to learn languages more implicitly (DeKeyser, 2003; DeKeyser & Larson-Hall, 2005; Lenneberg, 1967). However it is important to note that within the classroom context in many primary schools in the UK, learners may not be able to capitalise on their ability to learn implicitly, since exposure to the target language is substantially limited (DeKeyser, 2003; DeKeyser & Larson-Hall, 2005; Gass & Selinker, 2008; Muñoz, 2006, 2008b). Consequently this study will investigate the extent to which the explicit teaching of foreign language grammar can be effective and useful for primary school-aged learners learning German as a foreign language (RQ1).

There continues to be extensive debate as to whether explicit instruction simply results in the learner developing explicit knowledge of grammatical rules (e.g. Hulstijn, 2002; Krashen, 1982; M. Paradis, 1994) or whether it can also impact the learners' implicit knowledge and underlying grammatical system (e.g. DeKeyser, 1998, 2007; N. Ellis, 2005; VanPatten, 2002). In addition a related methodological issue plaguing effect of instruction research is the appropriateness of the chosen outcome measures for eliciting more explicit knowledge (e.g. knowledge of rules) and more implicit knowledge (e.g. ability to use in spontaneous discourse) respectively. Indeed it has been argued that the tasks (e.g. sentence-level, written, untimed) used in many studies in this area have tended to favour the use of explicit rather than implicit knowledge. Based on such tasks it is not possible to make claims about the impact of instruction on learners' more implicit knowledge and underlying grammatical competence (Doughty, 2003; R. Ellis, 2005, 2009b). In order to address this issue the present study will incorporate a battery of measures appropriate for testing not only the learners' comprehension and production of the target grammatical feature (in line with previous PI studies), but also measures which are thought to be more sensitive to eliciting different knowledge types available under different conditions (along the explicit-implicit continuum). The findings of this study will therefore contribute to the debate surrounding the type of knowledge developed following explicit grammar instruction (RQ2).

The instructional approach chosen is PI, which has been shown to improve learners' comprehension and production of a range of target grammatical features in numerous languages (e.g. Agiasophiti, 2013; Benati, 2001, 2005; Cadierno, 1995; Farley, 2004b; Laval, 2013; Marsden, 2006; Marsden & Chen, 2011; VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996; VanPatten & Wong, 2004). The present study will utilise two components of the PI package: explicit information and referential activities in which the target FMC is task-essential. To date only a handful of studies (Agiasophiti, 2013; Marsden, 2006; Marsden & Chen, 2011) have compared PI to an alternative form of input-based instruction, with Marsden's studies demonstrating substantially larger learning gains for PI over enriched input. The present study therefore aims to build on the findings of Marsden (2006) and Marsden and Chen (2011) by investigating whether making attention to the target grammatical form *only* (and not the relevant FMC) task-essential will result in equivalent gains to that of PI. The two interventions in the present study are labelled: task-essential form-meaning connection (TE-FMC); and task-essential form only (TE-F). Drawing such a comparison will test the fundamental claim of PI and Input Processing theory that attention to the FMC is necessary in order for the relevant grammatical form to be processed from the input (Marsden & Chen, 2011) (RQ3).

In line with the aims set out above, the present study will seek to answer the following research questions:

1) Does explicit grammar instruction improve young learners' a) comprehension and b) production of the target grammatical feature?

2) To what extent does explicit grammar instruction develop different types of knowledge of the target grammatical feature?

3) Following explicit information, is intentional practice in attending to the target form-meaning connection more beneficial than intentional practice in attending to the target grammatical form only?

Chapter 3: Methodology and Methods

This chapter discusses the methodology and describes the methods of the current study. Details regarding the experimental procedure, design and implementation of the activities, and statistical analysis are provided in the following sections, as well as information about the participants and ethical considerations of classroom-based experimental research.

3.1 Classroom-based experimental research

3.1.1 Classroom- versus laboratory-based experimental studies

Laboratory-based experimental research has traditionally been viewed as more robust than classroom-based experimental research (Hulstijn, 1997, p. 319; Mackey & Gass, 2005), due to the fact that in a laboratory context the researcher is more readily able to tightly control and manipulate experimental variables (e.g. random assignment of participants to treatment groups, amount of target language input the learner is exposed to etc) (Hulstijn, 1997; Mackey & Gass, 2005). In contrast, it has been argued that in the classroom environment it is difficult to control such intervening variables, thereby resulting in poor validity as well as difficulty in discerning the relationship between experimental variables and a lack of definitive causal claims (Hulstijn, 1997; Mackey & Gass, 2005). However whilst allowing the controlled investigation of, for example, language acquisition processes, the very nature of laboratory research, being abstract from real life, can lead to difficulties in extrapolating the findings to real life teaching and learning (Hulstijn, 1997; Schmidt, 1994). Consequently ecological validity is a particularly pertinent consideration when carrying out studies designed to test the effectiveness of a given teaching approach. Potentially, laboratory versus classroom-based studies may produce conflicting results (Spada, 2005); differences could be rooted in the fact that within the laboratory setting, a specific phenomenon is studied in isolation and target language input is strictly controlled, whereas within the classroom there are a multitude of factors (e.g. exposure to other types of linguistic input and other types of interaction), which could influence the way in which participants respond to a particular treatment (L. Cohen, Manion, Morrison, & Bell, 2011; Spada, 2005). Therefore, if the findings of effect of instruction studies are to have greater scope for informing classroom practice, they need to be carried out within the classroom context (Hulstijn & de Graaf, 1994; Nunan, 1991). For this reason, the present study was conducted within the classroom environment. The following sections consider the key characteristics of experimental research and potential threats to internal and external validity, which are synonymous with such research, particularly that which is carried out within the classroom context.

3.1.2 Experimental research

3.1.2.1 Characteristics of a formal experiment

Experimental research can be defined as “research in which variables are manipulated and their effects upon other variables observed” (Campbell & Stanley, 1966, p. 1). Within the educational and second language research contexts the manipulated variable (i.e. independent variable) often takes the form of an instructional “treatment” (Mackey & Gass, 2005, p. 137). The aim of experimental research, then, is to determine *causality*, in other words whether there is a causal relationship between the treatment (independent variable) and the participants’ performance on one or more outcome measures (dependent variable). A range of experimental designs can be employed depending on the nature and number of the chosen independent and dependent variable(s). The present study utilises a *between-group* (TE-FMC, TE-F, Control) *pre-, post-, delayed post-test* design, in order to determine whether the same gains (or lack of) were made following different (or no) treatments. The outcome measures employed in experimental studies often (although not always) yield quantitative (i.e. numerical) data. Therefore statistical analysis tends to be employed in order to determine the size and nature of the relationship between the independent and dependent variable(s) in question (Campbell & Stanley, 1966; L. Cohen et al., 2011; Mackey & Gass, 2005). The next section explores in more detail the key characteristics of experimental research, in relation to potential threats to internal and external validity.

3.1.2.2 Threats to validity

Given the drive to determine causality (i.e. What effect does X have on Y?), *control* constitutes a key issue in experimental research; “if rival causes or explanations can be eliminated from a study then clear causality can be established; the model can *explain* outcomes” (L. Cohen et al., 2011). Indeed experimental research carried out within the educational context can be notoriously complex due to the experimenters “lack of complete control” (Campbell & Stanley, 1966, p. 1). Nevertheless when carrying out an experimental study the researcher endeavours to control any extraneous or potentially intervening variables as tightly as possible in order to maintain the internal and external validity of the study. *Internal validity* can be defined as “the extent to which the results of the study are a function of the factor that is intended by the researcher” (Mackey & Gass, 2005, p. 358), in other words, are the observed changes in the dependent variable due to the experimental treatment (i.e. independent variable)? *External* *validity* relates to the generalizability of the findings to other populations, settings, treatments etc (Campbell & Stanley, 1966; L. Cohen et al., 2011; Mackey & Gass, 2005).

i) History, maturation, test effect

A number of factors can threaten the validity of a study, for example: *history*, the influence of events, which occur in addition to the treatment and between different measurement points; *maturation*, change within the respondents over time (e.g. age etc); *test effect*, responding to a measure at pre-test may affect responses at post-test (Campbell & Stanley, 1966; L. Cohen et al., 2011; Mackey & Gass, 2005; Torgerson & Torgerson, 2001). Randomisation has been identified as an optimal means of controlling such factors and ensuring the equivalence of the experimental and control groups, since any peripheral characteristics of the participant or environment, which may affect the outcomes of the study, are likely to be equally distributed across both groups. It is important to note that the tendency within a majority of classroom-based experimental research is to utilise a quasi-experimental approach whereby intact classes are assigned to either the experimental or control groups (Norris & Ortega, 2000; Spada, 2005). This is often deemed necessary due to the fact that random assignment of participants to different experimental groups, *within* one class is often not practicable (L. Cohen et al., 2011; Marsden & Torgerson, 2012; Spada, 2005). Further it could be argued that in certain cases (e.g. exploring the effects of a particular instructional method) an intact classroom may in fact be the most 'ecologically sound' context for the research study to take place in (Mackey & Gass, 2005, p. 143; Spada, 2005). Nevertheless the use of intact classes can make it difficult to control for extraneous variables (such as those above) which could potentially impact the findings of the study and make interpretation of the results problematic. Consequently, in the present study, the participants from four classes within two schools were assigned to either the TE-FMC or TE-F intervention group by matched randomization and additionally three classes were recruited from two schools to form a control group (section 3.3).

ii) Researcher as teacher

An important factor to consider is the influence of the person delivering the treatment and / or outcome measures. On the one hand the presence of the researcher, as a 'foreign body' in the classroom can produce a *Hawthorne effect* and alter the way in which the participants respond to the instructional and test materials (L. Cohen et al., 2011; Mackey & Gass, 2005; Torgerson & Torgerson, 2001). Conversely, utilising the class teacher to deliver the instructional materials to their own classes can be problematic, as variation can occur in the way in which the materials are delivered (*fidelity to condition*) and/or confounding variables, such as the class dynamics with particular teachers can potentially lead to differences in the way in which individual classes respond to an intervention (Spada, 2005). Furthermore, variation can occur between different classes in terms of the nature and amount of teaching the participants receive prior to, as well as during, lessons that occur simultaneously to the experimental study (Marsden, 2006; Spada, 2005; Spada & Lightbown, 1993). With regard to the present study, in order to address these concerns, the researcher taught German to all participant classes throughout the academic year in which the study took place (section 3.3.3) and delivered all of the intervention and test materials to all of the classes. Consequently, by the time the study took place the researcher was already a familiar presence in the classroom as the language teacher and could ensure that the materials were delivered systematically and reliably across the classes.

3.1.3 Ethical considerations and how they were addressed in the present study

The primary ethical issue within the present study related to the fact that a) two different teaching approaches were utilised and b) treatment was withheld from the non-active Control group. It is therefore important to consider the ethics of offering one intervention to some learners and a different (or no) intervention to other learners within the same class (Marsden, 2007). On the one hand it could be argued that withholding a treatment from some learners is unethical, since it may infer particular benefits that some learners are being denied. On the other hand, very little experimental research has been carried out within the primary school context in England; therefore it is not possible to definitively know the potential impact of the chosen instructional technique(s) before the study takes place. Nevertheless both of the treatments utilised in the present study were based on instruction which previous research (for reviews, see Norris & Ortega, 2000; Spada & Tomita, 2010; VanPatten, 2004c) has shown to result in at least some improvement in terms of learners' grammatical knowledge and / or use of the target language. In addition, whilst the non-active Control group did not receive any instruction as part of the intervention, they were given extra vocabulary training during the five week intervention period. Further following the completion of the study all of the schools and classes were given the intervention materials to use at their discretion.

It is also important to consider that for schools and practitioners, participation within research studies requires that some benefit for practice is perceived (Spada, 2005, p. 334). Notably the recruitment of schools to participate in the present study coincided with the government's announcement that from September 2014 foreign languages would be introduced as a compulsory foundation subject in primary schools. Therefore the present study was of particular relevance for practice, regarding how foreign language grammar can be taught within the primary classroom. Further, one issue with the implementation of compulsory foreign language teaching at primary school, which has been highlighted repeatedly (e.g. Board & Tinsley, 2014; Marilyn Hunt et al., 2005; Mitchell, 2011; Wade et al., 2009), is the lack of staff with relevant subject expertise. As part of the study, however, the researcher offered to spend one full academic year (voluntarily) teaching weekly German lessons to the Year 5 and 6 classes within the participating schools. This constituted a clear benefit for the participating schools.

It is also important to be wary of the extent of disruption which may be caused to the learners (Mackey & Gass, 2005). Within the present study the intervention sessions, as well as the written outcome measures, were incorporated into the learners' regular German lessons, thereby minimizing any potential disruption. The one-to-one outcome measures, however, were completed with each learner individually. During the pre-, post-, and delayed post-test weeks the researcher spent two days in school completing the activities with the learners in each class. This resulted in each learner missing approximately 15 to 20 minutes, of another lesson. This format was agreed with the class and head teacher prior to the study commencing. Further the learners' school commitments took precedence over participation in the research activities. If a participant was unable to participate in the activities at the allocated time (e.g. due to one-to-one numeracy tuition or a school trip), the researcher endeavoured to find an alternative time at which to complete the research activities with that learner.

As with any research study it was also important to ensure that the appropriate informed consent has been elicited prior to the study taking place. With regard to classroom-based research, this requires not only asking for consent from the participants, but also from other key stakeholders, such as the class teacher and head teacher (Mackey & Gass, 2005). In addition if the participants are children it is also important to inform parents as to the aims of the study. With regard to the present study, all interested parties were informed as to the general aims of the study, the procedure and the data recording tools to be utilised (Dictaphone, video camera). Both the head, or deputy head, teacher and the class teacher at each school were provided with an overview of the study and signed a consent form (Appendix 2). In addition, the parents were informed of the study via a letter sent home from school (Appendix 3). The letter was either opt-in or opt-out, at each school's discretion. Schools 1 and 2 agreed to proceed with an opt-out letter, whereby parents were instructed to contact the class teacher if they did not wish their child to participate. No such responses were received from any of the parents. School 3 chose to send an opt-in letter; parents were asked to return a short form stating whether they agreed to their child participating in the study. A small number of participants in School 3 stated that they did not wish to participate in the study. These learners did not take part in the one-to-one activities. Due to the fact that the written tasks were completed during the regular German lesson, all of the learners within each class completed these activities, but the scores of those learners who had 'opted-out' were not included in the analysis. The participants themselves were also informed as to the general aims of the study at the start of the lesson in the first week of the study. In addition, verbal consent was elicited from each of the participants before completing the one-to-one activities, and each participant was asked whether they were happy for the tasks to be recorded using the video camera. One Control group participant from School 2 stated that they did not wish to be video recorded, therefore only the Dictaphone was used to record the learners' responses.

Finally all stakeholders were informed that the data collected during the study would be held securely and confidentially by the researcher. Only the researcher had access to the full raw data set, and the supervisor was the only other person to view the raw data (or subsets thereof). Since the researcher needed to track each individual's performance over the course of the study, the data could not be collected anonymously. Nevertheless, no identifying information was included within the present thesis (nor in other dissemination documents such as presentations).

3.2 Selecting schools

Those schools offering German in the local area (n = 7) were approached (some initially by the supervisor) and invited to take part in the present study. Five schools responded with interest and a face to face meeting was arranged with each in order to discuss in more detail what the study would involve and to ascertain more in depth information about each school. A final pool of three schools was selected based on the following criteria:

- Each school had received Good to Outstanding in their most recent Ofsted examination
- Over 90% of participants had attained Level 4 or above in the 2012 KS2 English and Mathematics SATs tests
- Each school had already or intended to introduce German teaching at KS2
- There were at least 20 participants in Year 5 and Year 6 respectively

3.3 Participants

The participants (N = 139) were primary school children from three local primary schools. School 1 and 2 were both single-form entry and the Year 5 (age 9 to 10) and Year 6 (age 10 to 11) classes from each school took part in the study. School 3 contained three mixed Year 5 / 6 classes. Two of these classes were chosen to take part in the study by the school (due to time constraints it was not possibly to work with all three classes) and received German teaching throughout the year, whilst the third class was taught Italian by another teacher at the school.

The participants were younger than those who have taken part in a majority of other PI studies.

The three schools recruited for the present study contained mixed ability classes. Working with children of all abilities was essential in order to maximise the external and ecological validity of the study. A majority of UK primary schools will tend to utilise mixed ability classes (although some setting does occur in the higher year groups for subjects such as Literacy and Mathematics) (Sukhnandan & Lee, 1998), therefore it is important to determine whether the teaching approaches being studied are effective with a range of ability levels.

All of the participants were L1 speakers of English and were learning German as a foreign language in school. The learners had received two terms of weekly German lessons, (50 minutes per lesson; section 3.3.3) prior to taking part in the study and can be classed as beginner learners of German given this limited exposure to the language (Norris & Ortega, 2000, p. 454).

3.3.1 Matched pair randomisation

Matched pair randomisation was used within each class in schools 1 and 2 to assign participants to either the TE-FMC or TE-F intervention group. Matched pair randomisation involves matching one member of the experimental group with one member of the control or comparison group based on a relevant independent variable (L. Cohen et al., 2011, p. 319). In this case the independent variable was each participant's composite pre-test score on the two written outcome measures. Following the written pre-test, the participants within each class were ordered from highest to lowest. The participant with the highest score was then randomly assigned[25] to either the TE-FMC or TE-F group, and the participant with the second highest score was then assigned to the opposite group. The participant with the third highest score was then randomly assigned to the TE-FMC or TE-F group and the participant with the fourth highest score assigned to the opposite group, and so on. Randomisation, therefore, occurred at the pair level (L. Cohen et al., 2011, p. 319). It should also be noted that the randomisation process was 'blind' (Marsden, 2007). Participants were told that they were being split into two smaller teaching groups, named 'Germany' and 'Austria' and would take turns to complete German activities on the school laptops (section 3.5.6). In this way it was possible to avoid any impression of preferential treatment for one group or the other. Figure 3.1 illustrates the distribution of participants across the experimental and control groups.

[25] Random assignment was achieved via an online 2-level decision generator; https://www.random.org/coins

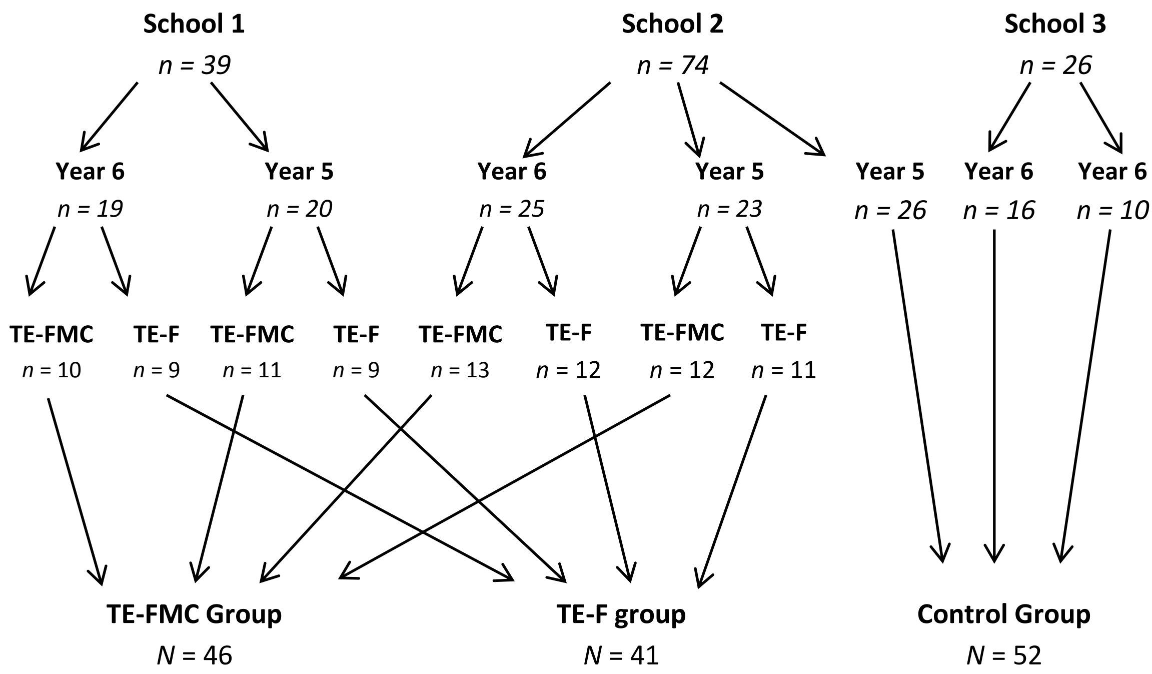


Figure 3.1: Distribution of participants across experimental groups

3.3.2 Control group

Due to the relatively small classes in Schools 1 and 2, it was not feasible to split each class three ways in order to also assign participants from within each class to the Control group. In addition the TE-FMC and TE-F groups received their respective interventions simultaneously during the regular German lesson, therefore, it would not have been possible to exclude any Control group participants from these lessons as the risk of 'contamination' would have been high. Consequently the Control group was recruited from one intact Year 5 class in School 2[26] and two mixed Year 5 / 6 classes in School 3 (Figure 3.1). As such the Control group can be described as "non-equivalent" (L. Cohen et al., 2011, p. 323) due to the fact that the participants were not chosen following randomisation. Nevertheless steps were taken in order to ensure the equivalence of the control and experimental groups. Firstly, approximately half of the Control group participants came from the same population as a sample of the TE-FMC and TE-F groups (School 2) (Kerlinger, 1970, cited in L. Cohen et al., 2011, p. 323). Secondly, School 3 had been matched on a number of factors with Schools 1 and 2 (section 3.2). Thirdly, the three groups (TE-FMC, TE-F, Control) completed pre-tests and a vocabulary test (section 3.6.6) in order to ensure that all of the participants had a similar level of baseline knowledge of the target feature and the relevant German vocabulary prior to the intervention (L. Cohen et al., 2011; Mackey & Gass, 2005; Norris & Ortega, 2000) (see section 4.2).

[26] The study was conducted with the TE-FMC and TE-F groups from schools 1 and 2 during the academic year 2012 to 2013. The Control group completed the study during the subsequent academic year, 2013 to 2014, which made it possible to recruit a second Year 5 class from School 2.

3.3.3 Pre-teaching phase

As noted in chapter 1, there is still significant variation in the quantity and nature of language teaching provided in primary schools across England (Board & Tinsley, 2014). Indeed Schools 1 and 2 were already offering weekly or fortnightly German lessons to Upper KS2 (Years 5 and 6) prior to participating in the current study, whereas language teaching in school 3 had been less structured with some French being taught in addition to a small amount of German provided by a teacher from the local feeder secondary school (the German teaching had not happened in the academic year prior to this study commencing). Furthermore although there are schemes of work available for primary languages (e.g. Cheater, 2007; DfES, 2005; QCA, 2007b), the content of language lessons can vary substantially between schools (Board & Tinsley, 2014). The content of language lessons in each of the participant schools was decided by the individual class teachers. Consequently the participants had differing levels of exposure to German, which varied both in terms of amount and content. This variability was a key concern for the present study, therefore it was decided that prior to the research study commencing, the researcher would teach German to the participant classes in all of the schools. This pre-teaching phase enabled a much greater degree of control regarding the language input the learners received and ensured that all participants across the three schools were familiar with the vocabulary (nouns, verbs) needed for the study.

The pre-teaching phase focused on providing learners with a core vocabulary and basic understanding of German. In line with the core topics covered in published primary-level schemes of work, topics such as greetings and introductions, family, pets and animals, hobbies, numbers, asking age, the classroom etc were taught as well as a core set of verbs and basic grammar (e.g. gender, subject/verb agreement) (Appendix 4). In accordance with the aims of the current primary foreign language curriculum, learners were given practice in listening, reading, speaking, and writing in the target language. Crucially, during the pre-teaching the learners received no exposure to the target feature (***den***).

At the time of being recruited for the study, the Year 6 classes from Schools 1 and 2 had already received a small amount of German teaching, amounting to approximately one term. Therefore, the start of the study was staggered for the Year 5 and 6 classes in these two schools. The Year 6 classes received one term of German teaching from the researcher and the Year 5 classes (along with the two classes from school 3) received two terms of German teaching with the researcher before taking part in the study. In this way, all participants had received a similar amount of German language instruction, including at least one term of German teaching from the researcher. In addition, the participants had minimal, if any, exposure to the target language outside of the classroom.

3.4 Design of the study

3.4.1 Experimental procedure

The study presented in this thesis was an experimental study consisting of four key stages:

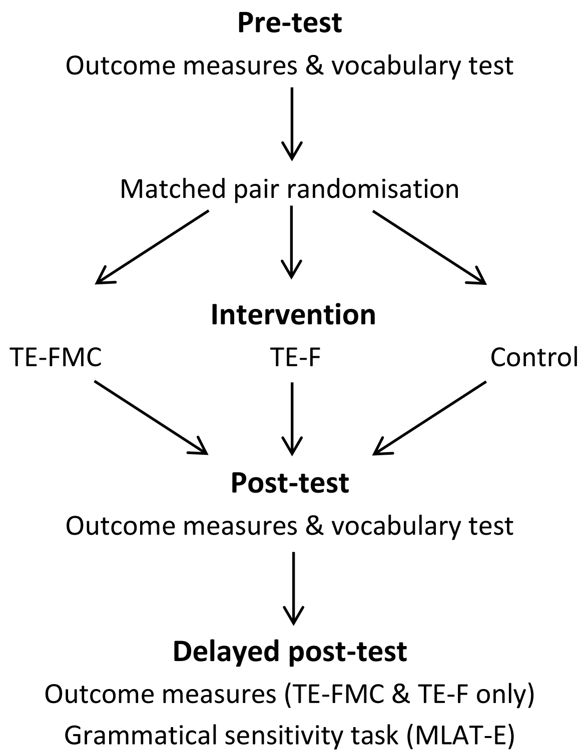


Figure 3.2: Experimental procedure

Table 3.1 details the time schedule of the study, which lasted for a total of 16 weeks:

Table 3.1: Time schedule of study

Week(s)	1	2 – 6	7	16
Stage	Pre-test	Intervention	Immediate post-test	Delayed post-test (TE-FMC, TE-F only)

The pre-test was administered one week prior to the intervention. The intervention itself was carried out over 5 weeks in weekly 50 minute sessions, giving a total duration of 4 hours and 10 minutes. Whilst this is acknowledged to be a relatively short intervention period, given the time constraint on access to the schools (one 50 minute lesson per week per class), as well as the occurrence of SATs test for the Year 6 class during the beginning of the summer term, it was thought to be realistic. Further, previous effect of instruction studies have used a similar length of instruction and 4 hours has been identified as a medium treatment length (Norris & Ortega, 2000; Spada & Tomita, 2010). The immediate post-test took place one week after the conclusion of the intervention. This was the earliest convenient time to carry out the post-tests in the participant schools following the conclusion of the intervention. Finally the delayed post-test took place 9 weeks after the post-test, in order to determine whether any improvement seen in the experimental group(s) was sustained after a substantial amount of time had passed (Mackey & Gass, 2005; Norris & Ortega, 2000). One possible issue with including additional post-tests after a long period of time is the risk of participants receiving additional exposure to the target feature in the interim (Mackey & Gass, 2005; Marsden, 2006). However, this issue was resolved by the researcher continuing as German teacher for all of the classes following the intervention, thereby ensuring that the learners received no extra exposure (intentional or incidental) to the target feature between the immediate and delayed post-tests.

The TE-FMC and TE-F groups completed the research study during the academic year 2012/2013, whereas the Control group completed the study in the academic year 2013/2014. Due to time constraints on access to the three Control group classes, it was not possible to carry out the delayed post-test with the Control group.

In order to limit the potential impact of any test effect occurring, due to the participants having completed the outcome measures three times (L. Cohen et al., 2011; Marsden & Torgerson, 2012), three versions (A, B, C) of each test were created (see section 3.6). The three versions were rotated between the year and experimental groups using a split block design:

Table 3.2: Rotation of outcome measure versions (A, B, C)

Group	*Age*	*Pre-test*	*Post-test*	*Delayed Post-test*
TE-FMC	**Year 6**	A	B	C
	Year 5	B	C	A
TE-F	**Year 6**	A	B	C
	Year 5	B	C	A
Control	**Year 6** **Year 5**	C	A	-

3.4.2 Piloting the intervention and test materials

3.4.2.1 Pilot study procedure

A pilot study was conducted in the academic year prior to the main study taking place. The aim of the pilot study was to determine the suitability of the intervention activities and outcome measures for the age group in question and identify and resolve any issues to do with their implementation and completion.

The Year 6 class (*N* = 27) from School 2 was recruited to participate in the pilot study, during the summer term of the academic year 2011 / 2012. The participants were therefore at an equivalent age (10 to 11) to the main study participants and had a similar level of German language knowledge, since they had received weekly 50 minute German lessons with the class teacher for approximately two terms prior to the pilot study taking place.

The pilot study was conducted over four weeks in the class' regular German language lesson. Table 3.3 details the pilot study procedure and the number of participants, who participated in each activity:

Table 3.3: Pilot study procedure

Week	Activity	*N*
1	What can you see	25
	Sentence Matching (*Pre-test*)	25
	Gap-fill (*Pre-test*)	25
2	TE-FMC	12
	TE-F	9
	Sentence Matching (*Post-test*)	21
	Gap-fill (*Post-test*)	21
	TE-F Think-aloud	6
3	Act-Out Comprehension	8
	Act-Out Production	8
	Sentence Reconstruction	8
4	Elicited Imitation	9

One question of particular importance related to the nature of the TE-F intervention. The TE-F intervention was designed as an extension to the enriched input approach which has been utilised in previous studies (e.g. Marsden, 2006; Marsden & Chen, 2011; Reinders & Ellis, 2009; J. White, 1998) by adding an extra noticing task in order to draw the learners' attention to the target grammatical form. In order to investigate the extent to which the TE-F intervention promoted such attention, six participants were recruited in the pilot study to complete the TE-F intervention as a think-aloud task.

In addition each of the proposed outcome measures[27] were piloted as well as one session of the proposed TE-FMC and TE-F intervention materials. A miniature pre- / post-test procedure was piloted using the written outcome measures. Due to the limited time available it was not possible to carry out a full pilot of the pre- and post-test procedure or the full intervention materials. After completing each activity the learners were asked to complete a brief feedback form relating to how easy and enjoyable the task had been. Notably, all of the tasks received overwhelmingly positive feedback; an important finding, since engaging learners with the task at hand is particularly crucial when working with young learners (section 3.5.1).

[27] Following the pilot study, the *What Can You See* task was removed from the study due to the complexity of the task, therefore will not be discussed.

3.4.2.2 Main findings and implications of the pilot study

i) Vocabulary familiarity

Lack of familiarity with the vocabulary used in the intervention activities and the outcome measures was found to affect the learners' performance when completing the tasks. A list of key nouns and verbs was provided to the learners with each of the pilot study tasks. In the feedback many reported relying on the vocabulary list to help them complete the activities. This issue was addressed for the main study by the pre-teaching phase in which all of the learners encountered the key vocabulary utilised in the main study activities.

ii) TE-F intervention

The think-aloud protocol carried out for the TE-F intervention revealed that many of the learners did not seem to be aware of the target feature with the input, despite completing the noticing task as part of the activity. Indeed many of the participants consistently read ***den*** (accusative article) as ***der*** (nominative counterpart), when discussing the target language sentences. Of course it should be noted that failure to mention the two different articles used within the target language sentences, does not necessarily mean that the learners had not noticed this difference and were not aware of it (Gass & Mackey, 2007; Hama & Leow, 2010). Nevertheless two features of the pilot version of the TE-F intervention are important to note. Firstly, both groups received information about the target grammatical feature, however only the TE-FMC learners received information about the potential processing problem (i.e. over-reliance on word order) and saw examples of OVS sentences containing the target feature. Withholding the examples of OVS sentences from the TE-F learners was motivated in line with the aim of the TE-F intervention, namely to draw the learners' attention to the target grammatical form, but not its meaning within the input. Nevertheless this difference may have served to disadvantage the TE-F learners and make isolation of the noticing practice impossible in drawing causal claims. Secondly, within the pilot version of the TE-F intervention activities, the learners completed the noticing task before completing the enriched input task; however this may have resulted in the learners mechanically completing the noticing task before engaging with the input in a meaningful way. Therefore for the main study, the same pre-practice explicit information was provided to both groups, and within the TE-F activities the learners engaged in the enriched input task before completing the noticing task for each item within the activities.

iii) Delivering the interventions

In the pilot study the intervention activity stimuli were presented via a computer, however the learners had to mark their answers on a worksheet. This meant that individual feedback could not be given to each learner based on their responses. It has previously been proposed that receiving correct/incorrect feedback can result in the learner inducing the target grammatical rule (e.g. DeKeyser et al., 2002; Sanz & Morgan-Short, 2004). Therefore for the main study, computer-based activities were developed in which the learners responded on the computer and were provided feedback after each activity item. For the pen-and-paper worksheet activities the learners also received feedback via answer sheets handed out after each activity.

iv) Outcome measures

With regard to the outcome measures, the pilot study revealed the importance of controlling animacy within the individual activities (see section 3.6.2), as well as the need to include feminine and neuter nouns within the written task stimuli in order to test the learners' ability to generalise the target rule (see section 3.6.3).

The other main issue which arose related to the pilot study *Elicited Imitation* task. The pilot version of this task was modelled on that of Myles and Mitchell (2012) and required the learners to produce target language sentences relating to a series of pictures. After every two sentences the learners were asked a comprehension question in English, which was designed to keep the learners focussed on the meaning of the target language sentences. In addition, the sentence stimuli contained both grammatical and ungrammatical sentences, in order to ensure that the task was reconstructive rather than relying on rote repetition (Erlam, 2006, 2009). However, this task proved problematic. Firstly, the meaning of each picture was not always clear to the learners. Secondly, the learners relied primarily on the picture when answering the comprehension question rather than the verbal test items (so meaning was not in focus). Thirdly, the learners demonstrated a tendency to overuse the nominative article (***der***) to the exclusion of the other possible articles, regardless of whether the sentence was grammatical or ungrammatical. This indicated that the learners were not relying on rote repetition when reproducing the sentences. Based on these findings, the task was redesigned for the main study (section 3.6.4.3).

3.5 The Intervention

3.5.1 Designing tasks for young learners

When designing the tasks to be used in the present study, one key factor to be considered was the age of the participants (9 to 11). In a study of young learners' perceptions of their language learning, Muñoz (2014) found that the learners, particularly those aged 11 years old, felt that they learned most from form-focussed activities (as well as vocabulary tasks). An examination of the learners' reasoning behind this choice suggested that this perception was due to the way in which such tasks explicitly focussed the learners' attention on a given grammatical form, which it was felt may in turn result in uptake. Notably, however, form-focussed activities were also identified by the learners' as one of the tasks they enjoyed least (Muñoz, 2014). It is therefore important to consider how such tasks should be designed in order to foster engagement and interest. Shak and Gardner (2008) carried out a study in which they sought to determine how young, primary school-aged language learners perceived four different types of focus-on-form tasks. Shak and Gardner (2008) observed three major influences on participant perceptions: production load, cognitive load, and pair or group work opportunities. The most positive reactions were given for those tasks which were cognitively stimulating, but not overly demanding and involved lesser production demands (Shak & Gardner, 2008). This is line with Hunt et al's (2005, p. 347) observation that 'children will only persist in learning tasks if they perceive them as worthwhile'. Further Harley (1998, p. 170) states that activities for primary school children should be stimulating and visually attractive. Noticing requires attention and for young learners, attention is dependent upon the intrinsic interest of the learning activities used. Consequently, the activities created for the present study were designed to incorporate visual stimuli, such as pictures and soft toys, in order to keep the learners engaged and attentive. In addition, the intervention activities consisted of computer-based activities, as well as pen-and-paper tasks, and a majority of the outcome measures (*Act-Out Comprehension*, *Act-Out Production*, *Sentence Repetition*, and *Sentence Reconstruction*) were interactive tasks completed on a one-to-one basis with the researcher.

3.5.2 Designing the intervention materials

All of the instructional materials were designed by the researcher, which ensured that the activities were age appropriate as well as suitable for the language ability of the participants. Accordingly, the intervention activities were built around the core vocabulary which the learners had encountered during the pre-teaching. Limited attentional resources are thought to be one cause of learners overlooking and not attending to grammatical form in the input (N. Ellis, 2006; Marsden, Altmann, & St Claire, 2013; Svalberg, 2012; VanPatten, 2004a, 2007). Therefore it was important to ensure that the participants were not pre-occupied with discerning the meaning of the lexical items within the input and could therefore focus on the target grammatical form (and its function). Both the TE-FMC and TE-F interventions consisted of two key elements:

i) explicit information about the target feature
ii) input-based interpretation activities

Each intervention group (TE-FMC, TE-F) completed three sessions of pen-and-paper worksheet activities and two sessions of computer-based tasks. The computer activities were created using Wondershare QuizCreator[28]. The pen-and-paper activities constituted reading tasks, whereas via the computer-based activities it was possible to also provide listening (as well as reading) activities for both groups. Since the intervention was delivered simultaneously for the TE-FMC and TE-F groups within each class it was not possible to include listening activities in the pen-and-paper tasks. Crucially, the TE-FMC and TE-F intervention activities contained the same (in nature and number) sentence stimuli (K = 212). All of the sentence stimuli were simple, transitive 'noun phrase – verb – noun phrase' (NVN) constructions. In weeks 1 to 3 the task stimuli contained only masculine (m) nouns (k = 120). In weeks 4 and 5 the learners also worked with sentences which contained one feminine (f) or neuter (n) noun as well as a masculine noun (m+f/n, k = 68; m+m, k = 24). This was important for the external validity of the intervention materials, since within the "real-world" (including the language classroom) one is likely to encounter input containing all three genders. In addition, the written outcome measures followed this format (m + f/n), therefore it was necessary to ensure that the learners had received some practice in interpreting input containing not only masculine nouns, but also feminine and neuter nouns. The sentences were presented in both SVO and OVS word order. Each activity contained an equal number of sentences in the two word order conditions (k = 106 for SVO and OVS respectively). Further it is important to note that the sentence stimuli utilised were identical across the TE-FMC and TE-F interventions (see sections 3.5.4 and 3.5.5).

3.5.3 Explicit information

The explicit information component of the instructional package was identical for both the TE-FMC and TE-F interventions. In line with the PI package (VanPatten, 2002; Wong, 2004a), it contained two core components:

i) information regarding the target grammatical feature and grammatical rules governing its use

ii) information regarding the processing problem which is often encountered by learners with regard to the target grammatical feature (overreliance on word order)

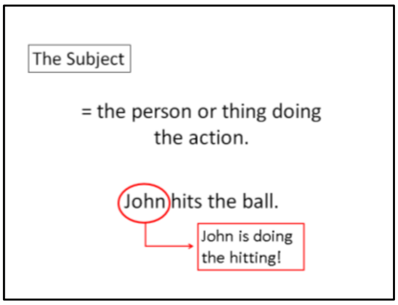
Over the course of the five week intervention, explicit information was provided four times, in weeks 1, 2, 4, and 5 (weeks 2 and 5 constituted a brief recap of the information given the previous week).

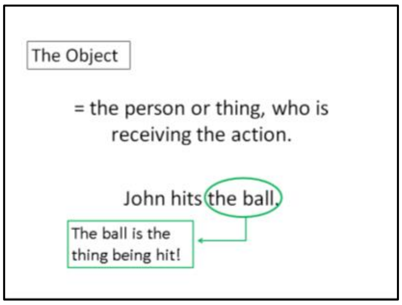
[28] http://www.wondershare.com/pro/quizcreator.html

Table 3.4: Amount of explicit information provided

Session	Length (approx.)	English examples	German examples SVO	German examples OVS
Week 1	15 minutes	3	3	3
Week 2	10 minutes	2	2	3
Week 3	-	-	-	-
Week 4	15 minutes	-	4	4
Week 5	10 minutes	-	2	2

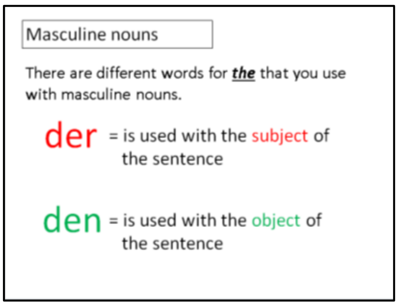
Given that the focus of the intervention was on the role of the masculine articles (***der*** and ***den***) in identifying the *subject* and *object* within German sentences, it was important to establish that the learners were comfortable with these terms in English (their L1), before proceeding with the explicit information and activities. Therefore in weeks 1 and 2 an additional element was included at the start of the explicit information, namely an explanation of the terms subject and object in English:

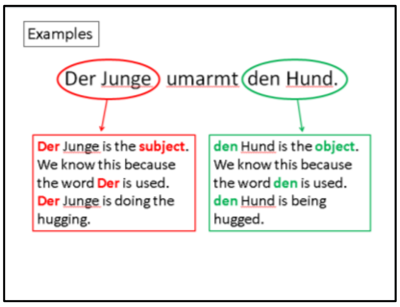




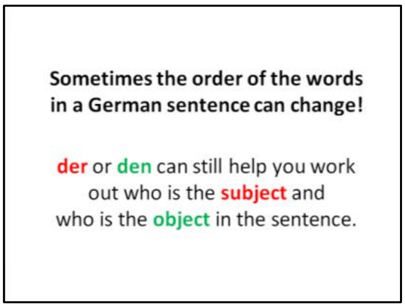
Example sentences (see Table 3.4) were included in order to give the learners practice in identifying the subject and object in English sentences before moving onto German.

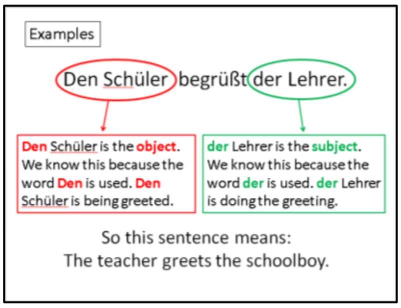
In weeks 1 and 2 the focus was on the two masculine definite articles, the accusative ***den***, and its nominative counterpart ***der***. The different functions of the two articles were explained to the participants and they were then shown examples of German sentences containing the two articles (see Table 3.4) and asked to identify the subject and the object:





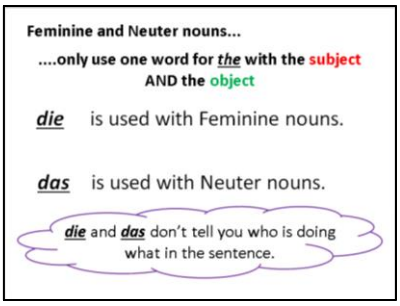
The importance of attending to the articles was then highlighted to the learners and it was explained that sometimes the order of words in German sentences can change:

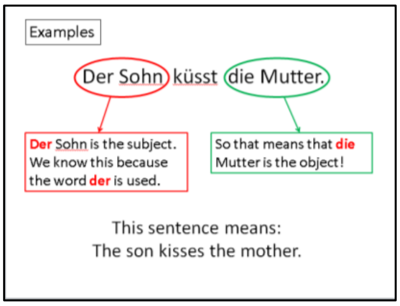




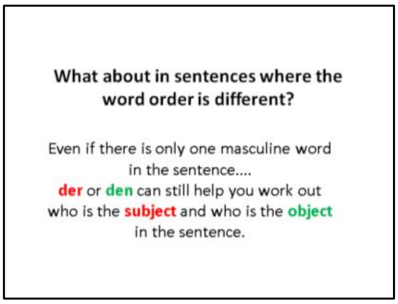
As above, the learners were shown examples of sentences in which the word order had been reversed (OVS) (see Table 3.4). The noun phrases were circled and annotated on screen in order to emphasise to the learners that they needed to pay attention to the articles in order to ascertain the grammatical role of the nouns within each sentence. In addition the learners were asked what the meaning of the sentence was in English, in order to ensure they had correctly understood the target grammatical rule.

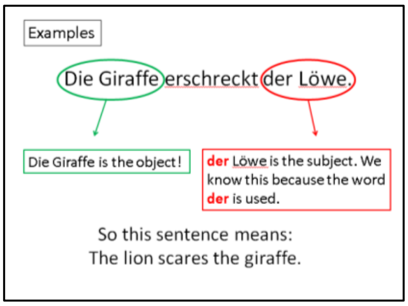
In weeks 4 and 5, the focus of the explicit information shifted to how ***der*** and ***den*** can be used to identify the grammatical role of the nouns within a sentence containing only one masculine plus one feminine or neuter noun. It was explained to the learners that with feminine and neuter nouns the same article is used for both the subject and the object of the sentence:





The learners were then shown examples of sentences in both SVO and OVS word order (see Table 3.4), in which the noun phrases were circled and annotated on screen, and it was reiterated that it is important to pay attention to the masculine article:





All of the explicit information was presented via PowerPoint and read out to the whole class (TE-FMC and TE-F learners simultaneously) by the researcher. Therefore the explicit information was identical for the two experimental groups and across all of the classes.

3.5.4 TE-FMC intervention activities

For the TE-FMC intervention, the aim was to design activities in which the learners were forced to attend to the target FMC in order to correctly complete the task and as such were akin to the referential structured input activities provided in PI (J. F. Lee & Benati, 2007; VanPatten, 2002; Wong, 2004a). Within such activities the input is manipulated so as to force the learners to be dependent on the target grammatical form in order to get meaning (VanPatten, 2002). With regard to the TE-FMC activities this 'input manipulation' involved removing the word order cue, on which L1 English learners are normally reliant, thereby pushing the learners to rely on the target feature (definite article case marking) in order to assign grammatical roles within the input. Accordingly the activities utilised SVO and OVS word order in equal number.

At the start of each activity session, the learners were given a brief recap of the target grammatical rule, either on the first page of the worksheet booklet or on the first few screens of the computer activity, which the learners could read through at their own speed (approximately one to two minutes). For the TE-FMC learners this included a reminder that learners have a tendency to rely on word order, but that when completing the activities they should pay attention to the different words for ***the***:

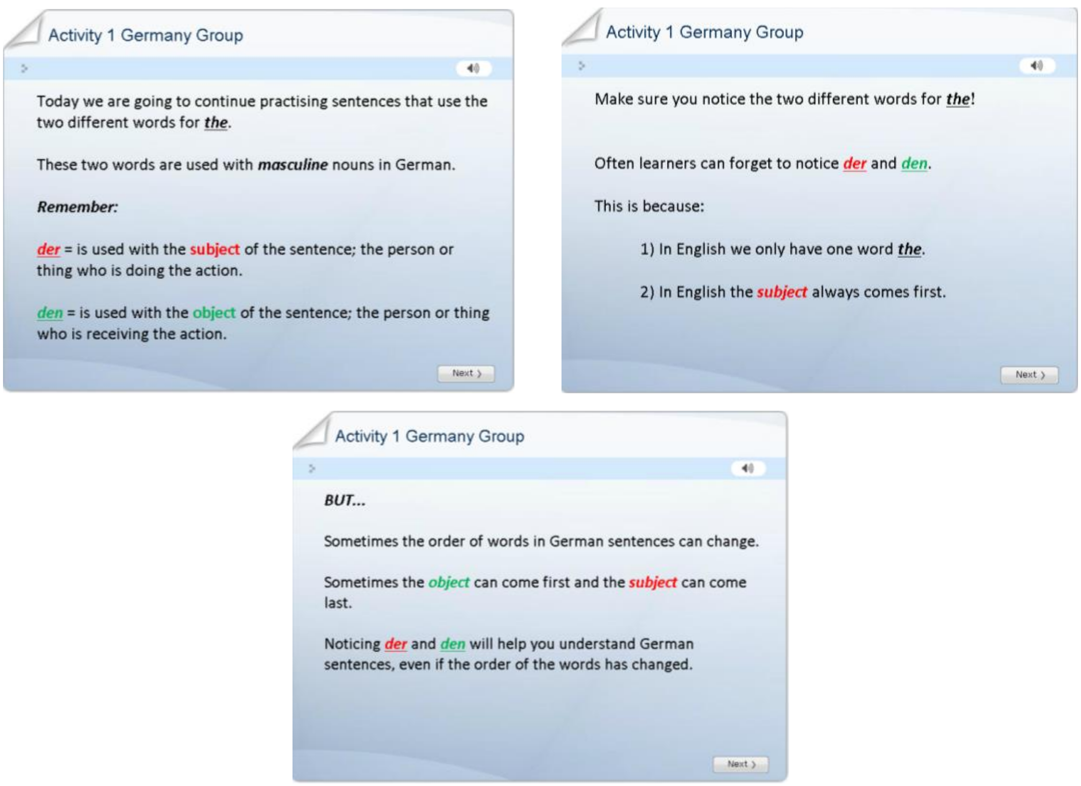


Figure 3.3: Recap of grammatical rule (TE-FMC)

Three activities were designed; two of which had listening and reading versions[29], giving a total of five activities which were rotated throughout the five weeks (Table 3.5). A variety of activities were included in order to maintain the learners' interest in the tasks. Each activity contained an equal number of items in SVO and OVS word orders. Examples of each reading activity are presented below (listening versions are provided in Appendix 5 and 6). The listening tasks followed the same format as each of the reading tasks; however the stimuli were presented aurally rather than in writing.

Table 3.5: Number of activities and items in intervention sessions (TE-FMC)

Activity	Modality	Gender	Week 1 Worksheet	Week 2 Computer	Week 3 Worksheet	Week 4 Computer	Week 5 Worksheet
Which picture?	Reading	m+m	16	12	-	-	4
		m+f/n	-	-	-	-	12
	Listening	m+m	-	12	-	8	-
		m+f/n	-	-	-	20	-
Who's doing what?	Reading	m+m	12	8	16	-	-
		m+f/n	-	-	-	8	-
	Listening	m+m	-	8	-	8	-
		m+f/n	-	-	-	12	-
Missing Nouns	Reading	m+m	20	-	16	-	4
		m+f/n	-	-	-	-	16
		Total items	**48**	**40**	**32**	**56**	**36**
		Total activities	**3**	**4**	**2**	**5**	**2**

3.5.4.1 *Which picture?* activity (TE-FMC)

The activities themselves (in particular the *Which picture?* activities) were designed in line with those of previous PI studies, in which the focus was also case-marking (e.g. Agiasophiti, 2013; Culman et al., 2009; Henry, Culman, & VanPatten, 2009; Stafford et al., 2012; VanPatten & Borst, 2012).

i) Activity

The *Which picture?* activity required the learners to read a target language sentence and decide which picture matched the sentence. Two pictures were presented for each item, and the pictures differed in who was performing the action (Figure 3.4). Due to the fact that the sentences were presented in either SVO or OVS word order (equal number of each) the learners were forced to attend to the target grammatical feature *and* its function in order to correctly assign grammatical roles within the sentence and subsequently select the correct picture.

[29] It was not possible to create a listening version of *Missing nouns* activity, since there was not an appropriate question format within the Wondershare QuizCreator software.

SVO sentence[30]: ***OVS sentence[31]:***

Figure 3.4: Items from Which picture? activity (TE-FMC)

In the computer-based activity (as above) the learners had to click on their chosen picture and then press submit. In the worksheet version of this activity the learners had to circle / tick their picture choice (Appendix 5).

ii) Feedback

For each activity the learners also received feedback, which ensured that the learners viewed attention to the target FMC as task-essential (Loschky & Bley-Vroman, 1993). Further it has been argued that feedback, even just correct/incorrect, can reinforce the target FMC thereby helping learners to induce the target grammatical rule, even when no pre-practice explicit information is provided (DeKeyser et al., 2002).

In the computer-based activities, the learners received feedback after each individual item, stating whether their chosen answer was correct or incorrect and presenting the sentence and the correct picture:

Correct answer (SVO): ***Incorrect answer (OVS):***

Figure 3.5a: Feedback for Which picture? activity (TE-FMC)

30 SVO: The-NOM lion asks the-ACC elephant.
31 OVS: The-ACC teacher greets the-NOM schoolboy.

If the answer for an OVS item was incorrect, the learners were reminded to pay attention to the different words for ***the***, but given no additional information about grammatical role assignment. For a sub-set of items (Table 3.6) the learners were also provided with the meaning of the sentence in English, for example:

Figure 3.5b: Extra feedback for Which picture? activity (TE-FMC)

Table 3.6: Number of items including extra feedback (Which picture? TE-FMC)

	Session	Total items	Extra feedback SVO	Extra feedback OVS
Computer	**Week 2**	24	4	4
	Week 4	28	4	4
Worksheet	**Week 1**	16	2	2
	Week 5	16	3	3

For the worksheet activities it was not possible to provide feedback after each individual item, since the TE-FMC and TE-F learners within a given class were completing the tasks simultaneously within one classroom. Therefore, after the learners had completed each activity they were given an answer sheet to mark their own answers. The answer sheet indicated the correct answer and for a subset of SVO items (Table 3.6) the learners were provided with the meaning of the sentence in English as above (Appendix 5). For a subset of OVS items (Table 3.6) the learners' attention was drawn to the target form and its function in the sentence:

OVS sentence:

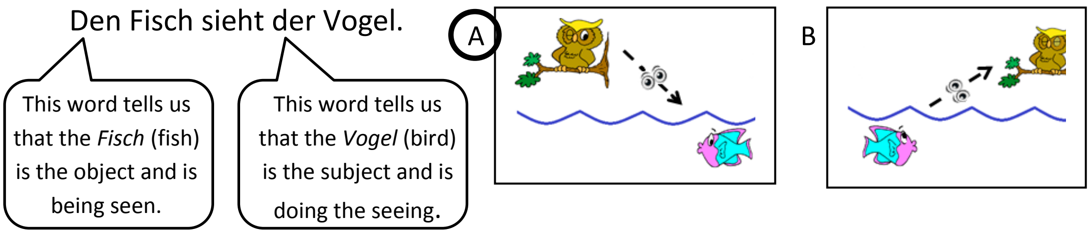


Figure 3.6: Feedback for OVS items in Which picture? worksheet (TE-FMC)

3.5.4.2 *Who's doing what?* activity (TE-FMC)

i) Activity

In the *Who's doing what?* task the learners had to answer questions based on a target language sentence, relating to who was doing or receiving the action. An equal number of SVO and OVS sentences were presented, therefore the learners had to attend to the target FMC in order to select the correct answer (see Appendix 6 for worksheet version), for example:

SVO sentence[32]: ***OVS sentence[33]:***

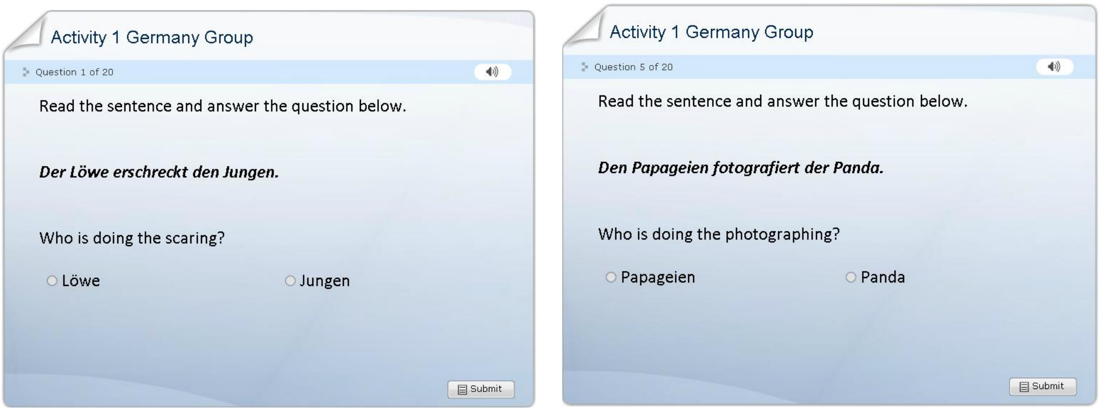


Figure 3.7: Items from Who's doing what? activity (TE-FMC)

ii) Feedback

The feedback for the computer-based version of this task confirmed whether the learner's answer was correct or incorrect, plus the target language sentence, and an explanation as to why:

Incorrect answer (SVO): ***Correct answer (OVS):***

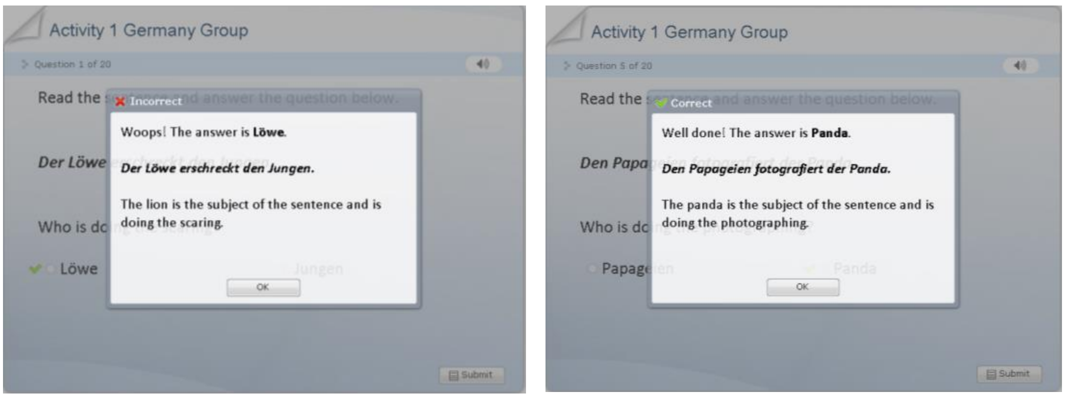


Figure 3.8: Feedback for Who's doing what? activity (TE-FMC)

The explanation was provided for all items and included explicit information identifying the subject or object within the sentence (in line with the question which had been asked for that

---

[32] SVO: The-NOM lion scares the-ACC boy.

[33] OVS: The-ACC parrot photographs the-NOM panda.

item). The answer sheet for the worksheet version of this task (Appendix 6) indicated the correct answer and an explanation (as above) was provided for a subset of the items:

***Table 3.7: Number of items including extra feedback** (Who's doing what? TE-FMC)*

	Session	Total items	Extra feedback	
			SVO	OVS
Worksheet	**Week 1**	12	2	2
	Week 3	16	2	2

3.5.4.3 *Missing nouns* activity (TE-FMC)

i) Activity

The *Missing nouns*, presented the participants with a sentence in which the two nouns had been removed. The learners were required to place each noun in the correct gap to make the sentence match the accompanying picture. The learners therefore had to attend to the target FMC in order to identify the correct position for each noun within the sentence (SVO or OVS in equal number), for example:

SVO sentence:

Der ______________ begrüβt den ______________.
[The-NOM ___________ greets the-ACC ___________ .]

Papagei [parrot] Löwen [lion]

OVS sentence:

Den ____________ erschreckt der ______________ .
[The-ACC ___________ scares the-NOM ___________ .]

Vogel [bird] Bär [bear]

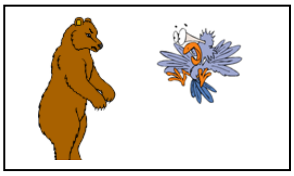


***Figure 3.9: Items from Missing Nouns activity** (TE-FMC)*

This task was included in the pen-and-paper activities only, since there was not an appropriate question format within the Wondershare QuizCreator software.

ii) Feedback

The answer sheet for the *Missing Nouns* activity detailed the correct position of the two nouns within each item. An explanation was provided for a subset of the items (Table 3.8), for example:

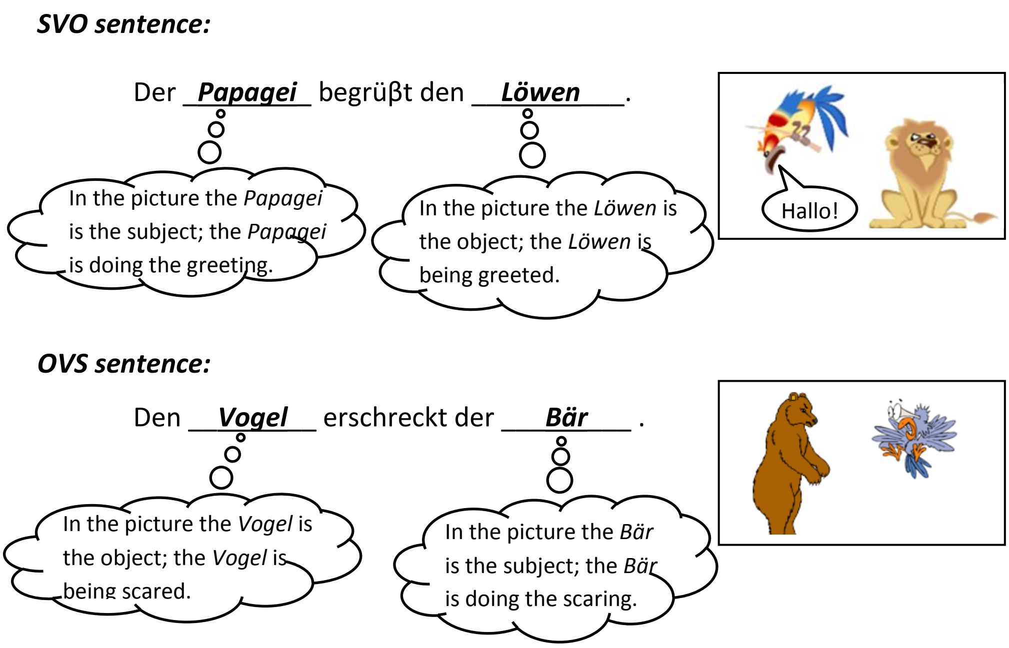


Figure 3.10: Feedback for Missing Nouns activity (TE-FMC)

Table 3.8: Number of items including extra feedback (Missing Nouns, TE-FMC)

	Session	Total items	Extra feedback	
			SVO	**OVS**
Worksheet	**Week 1**	20	3	3
	Week 5	20	4	4

3.5.5 TE-F intervention activities

The aim of the TE-F intervention activities was to draw the learners' attention to the target grammatical form only (i.e. not its meaning) within the input. For each activity, each item therefore consisted of two parts:

a) Enriched input tasks: focus on vocabulary practice

b) Noticing tasks: locating the grammatical form within the input

In line with previous studies which have utilised enriched input (e.g. Marsden, 2006; Marsden & Chen, 2011; Reinders & Ellis, 2009; J. White, 1998), the TE-F activities provided the learners with a flood of exemplars of the target feature; however the main focus of the activities themselves was on vocabulary practice. Unique to this study, however, an extra noticing task was incorporated into the activities, which asked the learners to identify the target feature within each activity item (section 3.5.5.2). As such this intervention constituted 'one step on' from an enriched input approach, since the learners' attention was directed towards the target feature. Further, in line with Svalberg's (2012) proposal, the learners were also required to engage with the target form. Notably, however it was 'one-step-behind' the TE-FMC activities, since the learners were not pushed to make the connection between the target grammatical form and its function within the input.

As with the TE-FMC intervention, the learners were given a brief recap of the grammatical rule before completing the activities. This was identical to the recap provided to the TE-FMC group, except that (given the aim of the TE-F intervention) the TE-F learners did not receive any extra information regarding the potential processing problem (an over-reliance on word order) at this point, although all of the learners were provided with this information during the pre-practice explicit information:

Figure 3.11: Recap of grammatical rule (TE-F)

Five TE-F activities were created. The noticing task (section 3.5.5.2) was incorporated into all of the computer-based activities, however was excluded from two of the worksheet activities in order to keep the activities varied and maintain the learners' interest (+ in Table 3.9):

Table 3.9: Number of activities and items in intervention sessions (TE-F)

Activity	Modality	Gender	Week 1 Worksheet	Week 2 Worksheet	Week 3 Computer	Week 4 Worksheet	Week 5 Computer
Picture Matching	Reading	m+m	16+	-	12	4	-
		m+f/n	-	-	-	12	-
	Listening	m+m	-	-	12	-	8
		m+f/n	-	-	-	-	20
Sensible or Silly?	Reading	m+m	12	16	8	-	-
		m+f/n	-	-	-	-	8
	Listening	m+m	-	-	8	-	8
		m+f/n	-	-	-	-	12
Missing Noun	Reading	m+m	20	16+	-	4	-
		m+f/n	-	-	-	16	-
		Total items	**32**	**16**	**40**	**36**	**56**
		Total activities	**3**	**2**	**4**	**2**	**5**

+Noticing task not included

It was crucial to minimise the differences between the TE-FMC and TE-F interventions, in order to limit the number of variables which may have affected the two groups' respective performances (L. Cohen et al., 2011). Therefore the TE-F activities were similar to the TE-FMC activities in terms of the mode of delivery (worksheet / computer-based), modality (written or aural stimuli), and whether pictures were included. In addition all of the TE-F activities utilised the **same** target language stimuli as the TE-FMC intervention, and therefore presented stimuli in both SVO and OVS word order (in equal number). It is important to acknowledge that OVS items would have offered the opportunity for the learners to induce the target FMC but, crucially, within the TE-F activities their attention was not explicitly directed to the target FMC. A description of the TE-F intervention reading activities is given below (listening versions are provided in Appendix 7 and 8).

3.5.5.1 *Picture Matching* task (TE-F)

i) Activity

In the *Picture matching* activity the learners read a sentence and were asked to decide whether it matched the picture. For those sentences which did not match the picture it was due to the fact that the verb stem was incorrect (i.e. due to lexical semantics):

SVO sentence[34]:

OVS sentence[35]:

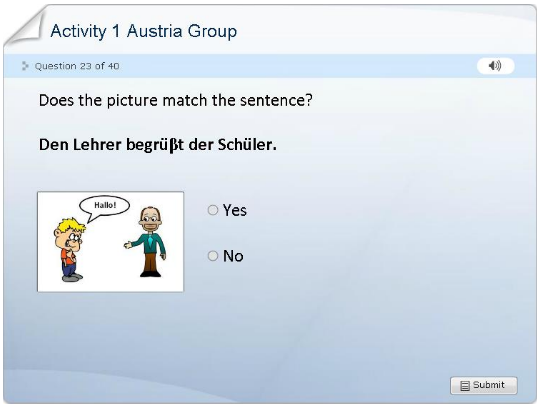


Figure 3.12: Items from Picture Matching activity (TE-F)

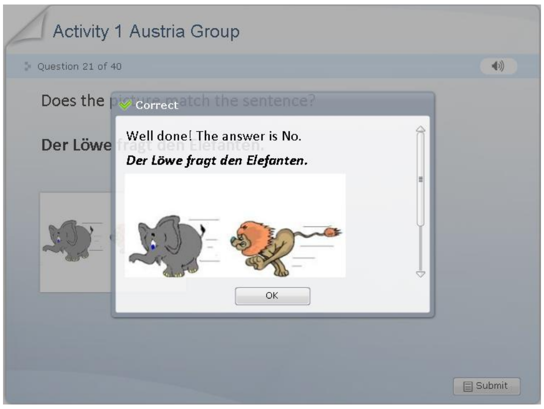
In the SVO example the sentence contains the verb *fragt* (asks), whereas the cartoon depicts the lion (*Löwe*) chasing the elephant (*Elefanten*). In contrast the OVS sentence contains the verb *begrüßt* (greets) and therefore does match the picture. By manipulating the verb stem in this way, the learners were primarily focused on the lexical items (i.e. verb stem) within each sentence, rather than the target FMC.

ii) Feedback

For the computer-based activities the learners received feedback after every item, which detailed whether they had chosen the correct or incorrect option, and provided the sentence and picture.

[34] SVO: The-NOM lion asks the-ACC elephant.
[35] OVS: The-ACC teacher greets the-NOM schoolboy.

Correct answer (SVO): ***Incorrect Answer (OVS):***



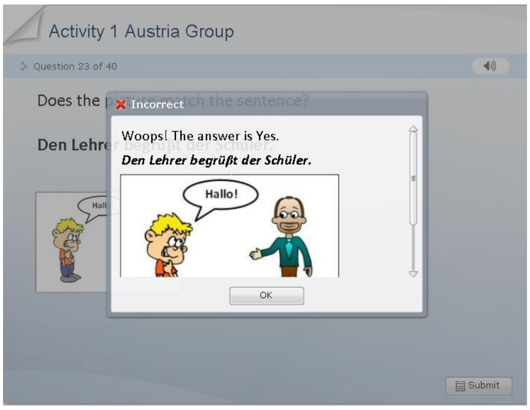


***Figure 3.13a: Feedback for Picture Matching activity** (TE-F)*

The sentence was provided in the TE-F feedback as this was the format followed in the TE-FMC activities. It was important to keep the level of exposure to the target feature within the input equal for both groups. For a sub-set of items (Table 3.10) the learners were also provided with an extra explanation as to why the picture did or did not match the sentence (i.e. the verb stem), for example:

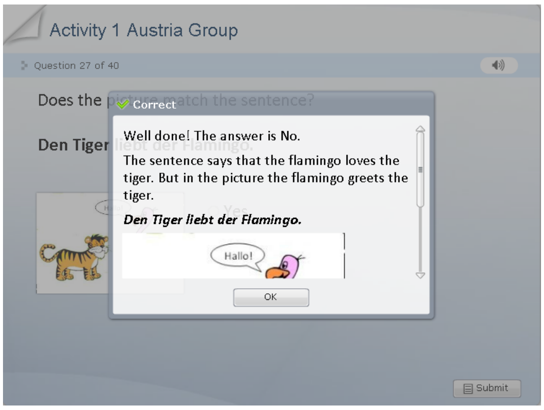


***Figure 3.13b: Extra feedback for Picture Matching activity** (TE-F)*

The answer sheet for the worksheet version (Appendix 7) indicted the correct answer and provided an explanation (as above) for a subset of items:

***Table 3.10: Number of items including extra feedback** (Picture Matching, TE-F)*

	Session	Total items	Extra feedback	
			SVO	OVS
Computer	**Week 2**	24	4	4
	Week 4	28	4	4
Worksheet	**Week 1**	16	2	2
	Week 5	16	3	3

It is important to note that whilst the content of the extra information within the feedback differed between the TE-FMC and TE-F activities respectively, this extra information was provided for the same sub-set of sentences within the computer-based and worksheet tasks.

3.5.5.2 Noticing task (TE-F)

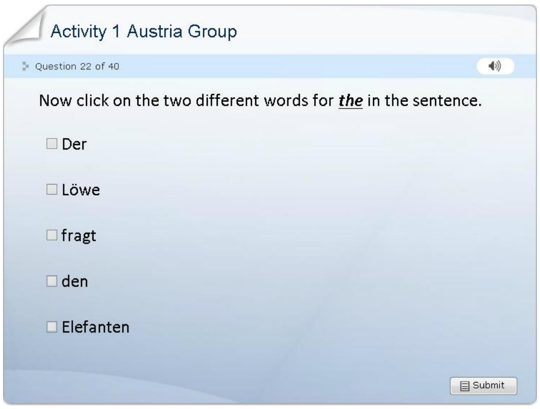
Within the TE-F enriched input activities (*Picture Matching*, section 3.5.5.1; *Sensible or silly?* section 3.5.5.3; *Missing Noun*, section 3.5.5.4) the learners' focus was on the meaning of the lexical items within the sentence and attention was not explicitly drawn to the target feature. However a second element was incorporated into all of the activities, the *Noticing task*, which asked the learners to identify the target feature (i.e. articles) within each sentence.

i) Noticing task (computer-based)

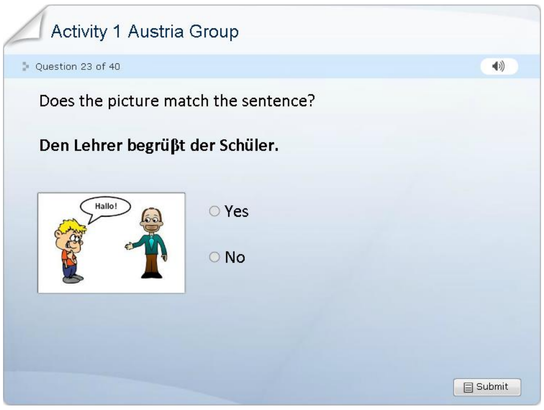
The learners completed the *Noticing task* immediately after they had completed the enriched input task for each item. The learners were asked to click on the words for ***the***:

Enriched input task (SVO):

Noticing task (SV0):



Enriched input task (OVS):



Noticing task (0VS):

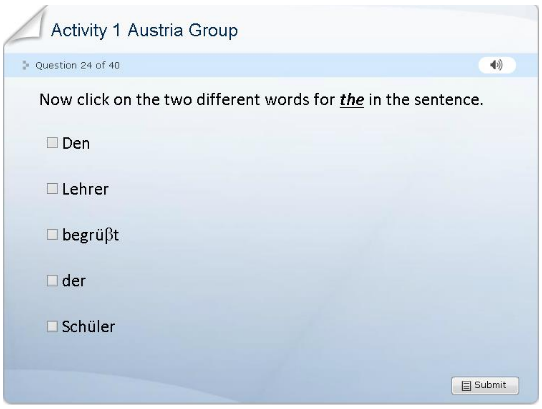


Figure 3.14: Noticing task items from Picture Matching task (TE-F)

The order in which the learner clicked on each word was not logged or made task essential. The learners completed the *Noticing task* for all of the items in all of the computer-based activities. This part of the task constituted the point at which the learners' attention was drawn to the target grammatical form (only). Crucially, however, the *Noticing task* did not make any explicit (i.e. task-essential on the part of the materials) connection between the target form and its function within the sentence stimuli. In the computer-based *Noticing task* the words within each sentence were listed vertically, with separate tick boxes, in order to avoid exposing the TE-F learners to the same sentence stimuli in the same format twice on two different screens within the activity.

Feedback for the Noticing task (computer-based)

If the learner correctly selected the two words for ***the*** within the sentence, then no feedback was provided aside from a message saying 'Well done'. If the learner selected the wrong word(s) then they were reminded of the correct words for ***the*** (the two articles were presented in isolation). It is important to note that there was a very low error rate on the *Noticing task* from the start of the intervention, therefore the incidence of the learners seeing the 'incorrect' feedback was very rare.

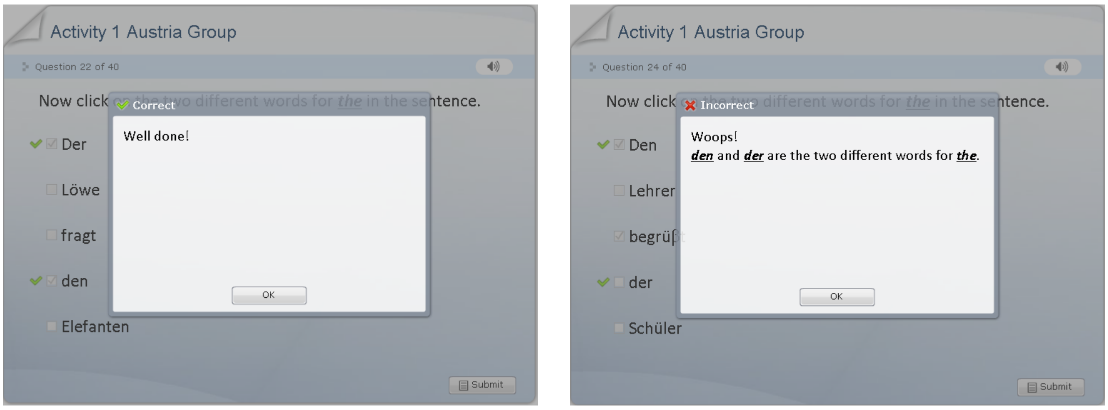


Figure 3.15: Feedback for computer-based Noticing task

ii) Noticing task (worksheet)

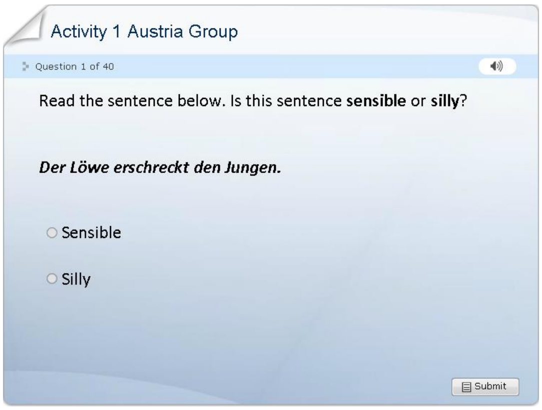
For the worksheet activities, the learners completed the *Noticing task* after they had completed the full enriched input activity, in order to avoid the tasks becoming too repetitive. At the end of a given task they were asked to look back at the sentences and circle the words for ***the***. Once both the enriched input activity and the *Noticing task* had been completed, the answer sheet was then provided. Within the answer sheet the correct words for ***the*** were circled (as well as the feedback relating to the enriched input activity). As detailed in section 3.5.5, the *Noticing task* was not included in the *Picture Matching* task in week 1 and the *Missing Noun* task in week 2.

3.5.5.3 *Sensible or silly?* task (TE-F)

i) Activity

In the *Sensible or silly?* task the TE-F learners were presented with a target language sentence and were asked to decide whether it was 'sensible' or 'silly' (see Appendix 8 for worksheet version):

SVO sentence[36]: ***OVS sentence[37]:***



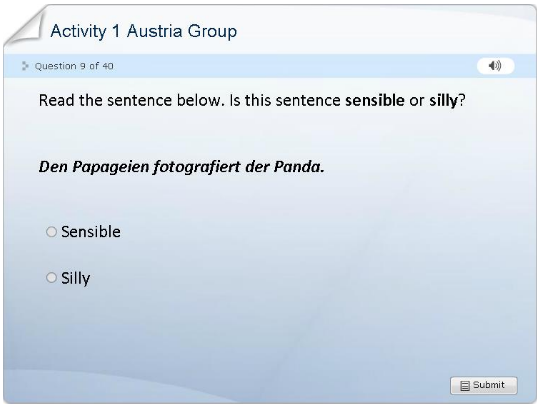


Figure 3.16: Items from Sensible or silly? activity (TE-F)

The decision as to whether a given sentence was 'sensible' or 'silly' related to the lexical semantics of the nouns and the verb within the sentence. However, it is important to note, that to a certain extent the interpretation of the sentence stimuli as 'sensible' or 'silly' was subjective. With the SVO example above, for instance, one learner remarked that the answer should be 'silly' *"because lions don't scare me!"*. Therefore, when completing the task, the learners were told to think about what might happen in general, but it was highlighted that it was okay to have a different opinion as the activity was just to provide practice in reading or listening to German sentences. The learners also completed the *Noticing task* within this activity (section 3.5.5.2).

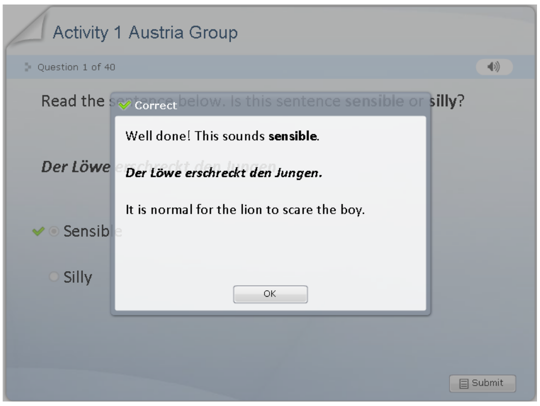
ii) Feedback

The feedback indicated whether the learners had answered correctly and provided the sentence stimuli, along with an explanation as to why the sentence might be thought of as 'sensible' or 'silly'. The explanation was provided for all of the items within the computer-based version of this task, in order to retain parity with the amount of feedback provided in the equivalent TE-FMC activity (*Who's doing what?*), for example:

---

[36] SVO: The-NOM scares the-ACC boy.
[37] OVS: The-ACC parrot photographs the-NOM panda.

Correct answer (SVO): ***Incorrect answer (OVS):***



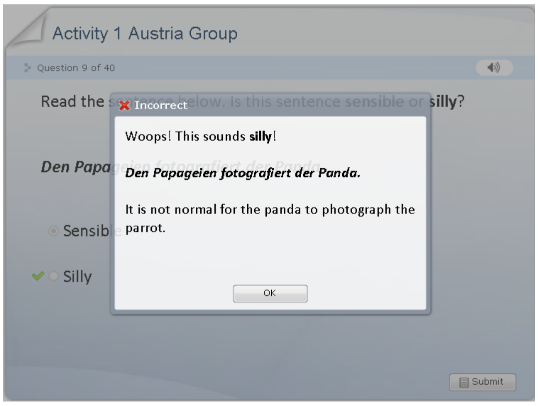


Figure 3.17: Feedback from Sensible or silly? activity (TE-F)

The primary focus of the explanation was on the affective judgement which had been made, and it should be reiterated that the learners were told that it was acceptable to have a different opinion about a given sentence. Nevertheless, it is important to acknowledge, that the explanation may have allowed the target FMC to be induced, given that it included a (paraphrased) translation of the sentence stimuli. The learners received the explanation (as above) for a subset of items in the answer sheet for the worksheet version of this task (Appendix 8):

Table 3.11: Number of items including extra feedback (Sensible or silly, TE-F)

	Session	Total items	Extra feedback	
			SVO	OVS
Worksheet	**Week 1**	12	2	2
	Week 3	16	2	2

3.5.5.4 *Missing Noun* activity (TE-F)

i) Activity

The TE-F *Missing Noun* activity was similar to the TE-FMC *Missing Nouns* task; however only one noun missing from the sentence:

SVO sentence:

Der _____________ begrüßt den Löwen.
[The-NOM ___________ greets the-ACC lion.]

Papagei [parrot] Schmetterling [butterfly]

OVS sentence:

Den ____________ erschreckt der Bär.
[The-ACC ___________ scares the-NOM bear.]

Affe [monkey] Vogel [bird]

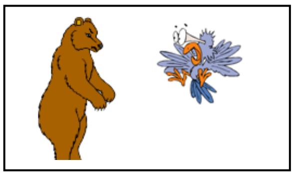


Figure 3.18: Items from Missing Noun activity (TE-F)

The learners were provided with a choice of two nouns which could fill the gap, based on the accompanying picture. The focus of this activity was on vocabulary practice, since the correct answer could be selected by determining which of the two nouns was depicted in the accompanying cartoon.

ii) Feedback (worksheet)

The answer sheet for the *Missing Noun* task provided the learners with the correct noun in each sentence. For a subset of items (Table 3.12) the learners were also provided with extra information which detailed the meaning of the two nouns, for example:

SVO sentence:

Der ***Papagei*** begrüßt den Löwen.

The missing word is ***Papagei***, which means ***parrot***. ***Schmetterling*** is the German word for ***butterfly.***

OVS sentence:

Den ***Vogel*** erschreckt der Bär.

The missing word is ***Vogel***, which means ***bird***. ***Affe*** is the German word for ***monkey.***

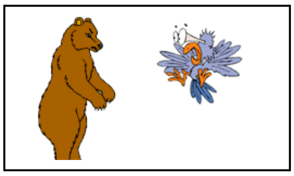


Figure 3.19: Feedback for Missing Noun activity (TE-F)

Table 3.12: Number of items including extra feedback (Missing Noun, TE-F)

	Session	Total items	Extra feedback	
			SVO	**OVS**
Worksheet	**Week 1**	20	3	3
	Week 5	20	4	4

3.5.6 Administering the intervention

The intervention was delivered to the TE-FMC and TE-F groups during each class' regular German lesson. The researcher, as the class German teacher, delivered all of the intervention sessions, thereby eliminating any possible 'teacher variable' (Marsden, 2007; Spada, 2005). The pre-practice explicit information was delivered to each class as a whole. Each group then completed their respective activities. In week 1 of the intervention both the TE-FMC and TE-F groups completed pen-and-paper worksheet activities, but in weeks 2 to 5 the groups alternated between pen-and-paper tasks and computer-based activities. The worksheet activities were distributed in individual booklets to the relevant group(s) following the explicit information. At the end of each activity the learners saw the following message:

STOP!
Mark your answers using the answer sheet.
Remember to use a different colour pen or pencil!

The answer sheets were distributed by the researcher as they were needed. The computer-based activities were stored in individual folders for each week on individual USB sticks and each activity within the set was numbered. The learners worked through the activities at their own pace.

It was anticipated that some of the learners within a given class were likely to work through the intervention activities more quickly than others. Therefore extra non-intervention activities were available once the intervention activities had been completed. These activities were based on topics that the learners had encountered during the pre-teaching phase (e.g. *Café, das Wetter*). Crucially the activities did not contain any instances of the target feature (***den***), thereby eliminating any potential for extra exposure. During the intervention sessions, the class teacher or a teaching assistant was present, to help with opening the computer-based activities, and distributing the answer sheets and non-intervention activities. However, they had no interaction with the participants about the German.

It has been argued that a key feature of experimental designs relates to ensuring that the experimental groups are kept entirely separate throughout the intervention in order to avoid any 'contamination' (L. Cohen et al., 2011, p. 313). However due to the fact that the intervention sessions took place during each class' regular German lessons this was not possible. Nevertheless steps were taken to ensure that the TE-FMC group was not exposed to the TE-F group activities and vice versa:

- the two groups were seated on opposite sides of the classroom
- there was minimal movement of participants around the classroom during the lessons
- in weeks 2 to 5, participants in one group completed listening activities on individual laptops with headphones whilst the other group completed pen-and-paper tasks
- feedback was delivered to participants individually (automatically in the computer-based activities / via answer sheets for the pen-and-paper activities)
- there was minimal group discussion of the activities and answers

The Control group in the present study was a 'non-active' control group and did not receive any teaching relating to the target feature during the intervention period (Campbell & Stanley, 1966; Marsden & Torgerson, 2012; Norris & Ortega, 2000; Torgerson & Torgerson, 2001). Rather the content of German lessons reverted back to the program of lessons being taught prior to the pre-tests. However, one important consideration which arose was the fact that the vocabulary used in the outcome measures was also used in the intervention materials for the TE-FMC and TE-F groups. Consequently one concern was that these groups would be at an advantage when completing the post-tests as they had received a substantial amount of extra exposure to the core vocabulary and lexical semantics of the core nouns and verbs. Therefore over two of the five lessons during the intervention period the Control group completed extra activities focussing on the core vocabulary used in the study (Appendix 9), in order to minimize any potential advantage, due to familiarity with the vocabulary, which the TE-FMC or TE-F group may have gained. Crucially, however, the activities contained no instances of the target feature ***den***.

3.6 Outcome measures

3.6.1 Measuring different types of knowledge

When assessing the effectiveness of a particular instructional approach, it is important to consider the nature of the tasks being used and the type of knowledge which they are likely to elicit (DeKeyser, 2003; Doughty, 2004; R. Ellis, 2005, 2009b; Han & Ellis, 1998). PI studies have traditionally included tasks to test learners' ability to a) interpret and b) produce the target feature (VanPatten, 2002). Nevertheless a shortcoming of a majority of effect of instruction research is that the types of tasks typically utilised are not necessarily appropriate for testing the development of knowledge which is useable under time and communicative pressure (Doughty, 2003, 2004; R. Ellis, 2005, 2009b; Norris & Ortega, 2000) (section 2.3.5). Rather the tasks often employed in such studies tend to be untimed and favour the use of metalinguistic, declarative knowledge about language (as noted by DeKeyser et al., 2002; Doughty, 2004; Marsden & Chen, 2011).

The present study sought to contribute to research in this area and to this end six outcome measures were designed in order to tap into different types of knowledge, which were broadly defined as explicit, implicit, and metalinguistic knowledge. The tasks were designed in accordance with the operationalization of these knowledge types, i.e. based on *time available*, *focus of attention*, and *metalanguage* (section 2.2.6). Han and Ellis (1998) argued that the type of knowledge employed in a particular task depends upon two key elements, accessibility and awareness. Implicit knowledge, for example, is accessed automatically in fluent spontaneous communication, whereas access to explicit knowledge requires 'controlled effort', and therefore is employed in tasks which require or allow for planning and monitoring (Han & Ellis, 1998). Further implicit knowledge is said to be acquired and held without awareness, in contrast to explicit knowledge which represents conscious, sometimes metalinguistic knowledge, of the language (Han & Ellis, 1998). Timed tasks, for example, tend to be completed based on 'feeling' or intuition' and are, therefore, often identified as measures of more implicit knowledge, since they do not require and further can constrain access to explicit knowledge (R. Ellis, 2005; Han & Ellis, 1998). Further oral discourse-level tasks are thought to be valid measures of implicit knowledge as they are 'online' meaning-focused tasks and do not provide much opportunity for planning or monitoring of language use if a communicative and interactive purpose is maintained (Doughty, 2004; R. Ellis, 2005). In contrast untimed tasks and tasks involving grammaticality judgements are more likely to tap into more explicit knowledge (R. Ellis, 2005). Notably, however, it is important to acknowledge the fact that learners are likely to rely on both types of knowledge to a certain extent when completing any given task (R. Ellis, 2002, 2005; Roehr, 2008). DeKeyser (2003) noted that "time pressure makes the use of explicit knowledge harder but does not exclude it completely" (p. 326), and the opposite (i.e. use of more implicit knowledge on untimed tasks) is also plausible. It is not, therefore, possible to definitively state that a given task will solely elicit one type of knowledge or the other. Nevertheless tasks should be designed to promote the use of a given type of knowledge to as great an extent as possible. Table 3.13 provides a summary of the outcome measures utilised in the present study:

Table 3.13: Outcome measures

	Comprehension	**Production**
Explicit knowledge (*written*)	Sentence Matching	Gap-fill
Implicit knowledge (*oral*)	Act-Out Comprehension	Act-Out Production
	Sentence Repetition	
Metalinguistic knowledge	Sentence Reconstruction	

In line with RQ1, the tasks were designed to test the learners' comprehension and production ability. With regard to RQ2, the two written tasks (*Sentence Matching*, *Gap-fill*) were untimed, sentence-level tasks and as such were designed to elicit more explicit knowledge of the target feature, since the learners would have the opportunity to monitor and reflect on their language comprehension and/or use. In contrast, the oral tasks (*Act-Out Comprehension*, *Act-Out Production*, *Sentence Repetition*) were completed one-to-one with the researcher and therefore exerted a much greater time and communicative pressure[38] than the written tasks. In addition the use of soft toys was incorporated into the oral tasks in order to keep the learners focussed on meaning. As such the oral tasks were designed to tap into more implicit

[38] Although the tasks themselves were also untimed

knowledge of the target feature. Notably the oral tasks still constituted sentence-level tests of comprehension and production; however, given the age and language ability of the participants discourse-level tasks were deemed inappropriate. Finally, a metalinguistic task was included in order to test the learners' ability to verbalise the target grammatical rule. Due to the fact that the learners were focused on form, were under no time pressure and were required to employ their metalinguistic knowledge this task constituted a clear test of explicit knowledge (R. Ellis, 2005).

3.6.2 Controlling animacy

The aim of the outcome measures was to test whether the learners were able to correctly comprehend and produce the target feature within both standard (SVO) and 'reversed' word order (OVS) sentences. Notably, however, when the test materials were piloted, it was found that on the pilot version of the *Act-Out Comprehension* task, some learners were able to correctly interpret an OVS sentence in which the subject was animate and the object inanimate[39], e.g. *Den Tisch putzt der Löwe* (the-ACC table cleans the-NOM lion). This finding indicated that some learners may have been relying on animacy cues as well as word order when interpreting the target language sentences. Indeed, previous research has demonstrated that although word order is the most reliable cue in English, animacy cues may also be utilised by L1 English speakers (Kempe & MacWhinney, 1998) and further may supersede word order cues when the two are in conflict (Jackson, 2007) (see section 2.3.3.3). Consequently for the main study, animacy was controlled across the written and oral outcome measures[40]. The activities contained an equal number of sentences in the following two conditions:

a) Animate subject + Animate object (A+A)
b) Animate subject + Inanimate object (A+I)

Additionally, the comprehension tasks (Sentence Matching, Act-Out Comprehension) contained an equal number of sentences in the following condition:

c) Inanimate subject + Animate object (I+A)

Due to the length and number of outcome measures to be used, it was not possible to include this third condition in the Gap-fill and Act-Out Production tasks. Controlling animacy in this way enabled an examination of whether the learners were relying on animacy in order to aid their interpretation of target language sentences. If the learners were relying on animacy, as well as word order, two hypotheses can be put forward regarding their performance:

[39] OVS sentences containing two animate nouns were interpreted incorrectly, indicating that the learners were not attending to the target FMC.
[40] The Sentence Repetition task contained only 6 items, therefore only the A+A animacy condition was included.

A) At pre-test the learners will have been able to correctly interpret OVS test items containing an Animate subject and Inanimate object
B) The largest gains between pre- and post-test will have been made on test items containing an Inanimate subject and Animate object, across both word order conditions

In order to explore these hypotheses, Animacy was isolated as a variable within the analysis[41], in order to investigate the extent to which the learners' performance across the Animacy conditions changed between pre- and post-test.

3.6.3 Written outcome measures

3.6.3.1 Designing the written activities

The written activities were pen-and-paper tasks, which the learners completed at their own pace. The test stimuli for the written activities were simple transitive NVN constructions and were generated from the list of core vocabulary (Appendix 4), which was deemed familiar to the learners following the pre-teaching phase. Cartoon pictures were generated, using ClipArt and Microsoft Windows Paint, to accompany each of the test stimuli. In line with the split-block design, three versions of each written task were created. Due to limitations on the range of cartoon images that could be created, it was not possible to generate novel test items for each version of the tasks. Therefore the same test stimuli were utilised in all versions; however the following steps were taken in order to ensure that each version was sufficiently different from the others:

- The word order of test items 1 to 12 from version A was reversed (i.e. SVO to OVS and vice versa) for version B
- The word order of test items 13 to 24 from version A was reversed (i.e. SVO to OVS and vice versa) for version C
- The test stimuli in versions B and C were reordered using the Randomise – RAND() – and SORT functions in Microsoft Excel

3.6.3.2 *Sentence Matching* task

The *Sentence Matching* task borrows elements from the Grammar Interpretation tasks used by Shak and Gardner (2008) and in line with previous PI studies (e.g.Benati, 2005; Marsden, 2006; VanPatten & Borst, 2012; VanPatten & Cadierno, 1993) tested the learners' ability to correctly comprehend the target feature. The learners were presented with sentences in which the target grammatical feature (case-marking) conflicted with the cue common to the learners' L1

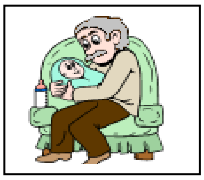
[41] The same pattern of results was found across all of the outcome measures; therefore only the results of the Act-Out Comprehension task are presented (see section 4.5).

(word order), thereby making it possible to determine which cue(s) the learners were relying on when interpreting the target language sentences (Gass & Mackey, 2007; Wong, 2004a). The participants were presented with pairs of sentences (*K* = 24) in either SVO (*k* = 12) or OVS (*k* = 12) word order plus a corresponding picture and were asked to decide which sentence (A or B) matches the picture.

SVO sentence pair:

A. Das Baby umarmt den Opa. A B
(*the-NOM/ACC baby hugs the-ACC grandfather*)

B. Der Opa umarmt das Baby.
(*the-NOM grandfather hugs the-NOM/ACC baby*)



OVS sentence pair:

A. Die Mutter verfolgt der Sohn. A B
(*the-NOM/ACC mother chases the-NOM son*)

B. Den Sohn verfolgt die Mutter.
(*the-ACC son chases the-NOM/ACC mother*)

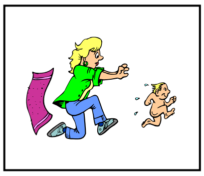


Figure 3.20: Items from Sentence Matching task (Set B)

Within each sentence pair the subject / object role assigned to each noun was reversed (e.g. *Baby* was the subject and *Opa* the object in sentence A; vice versa in B). In addition, each sentence contained one masculine noun and one feminine or neuter noun. As observed previously (section 2.5) the article used with feminine (*die*) and neuter (*das*) nouns respectively does not change between the nominative and accusative cases (Dreyer & Schmitt, 2001; Jackson, 2007), therefore in order to interpret each sentence correctly the learners would have to pay attention specifically to the article used with the masculine noun (with two masculine nouns it would not have been possible to isolate which article the learner was relying on). It is also important to note that across the SVO and OVS conditions there was an equal distribution of sentences containing the masculine noun in the subject and object position respectively.

It was hypothesised that the learners would perform at ceiling-level on the SVO sentences at pre-test, therefore, any learning gains observed at post- and delayed post-test would be due to an improvement in the participants' comprehension of the sentences in OVS word order. An improvement in the learners' interpretation of the OVS sentences would demonstrate that those learners were attending to the target feature (case-marking on masculine nouns) and 'overriding' the word order cue from their L1 (Jackson, 2007) and as predicted by the FNP (VanPatten, 1996, 2004a, 2007).

Scoring procedure

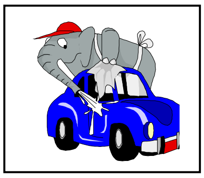
The total possible score on this task was 24. One point was awarded for each sentence correctly identified as matching the picture from each pair (SVO = 12, OVS = 12). All test papers were marked by the researcher.

3.6.3.3 *Gap-fill* task

The *Gap-fill* task presented the learners with a picture and a corresponding sentence in which the masculine noun phrase was missing (*K* = 24):

SVO sentence + missing subject (SVO+Subj)

…………………………… (the elephant) wäscht das Auto.
washes the-NOM/ACC car.



OVS sentence + missing subject (OVS+Subj)

Das Baby füttert …………………………… (the father).
the-NOM/ACC baby feeds

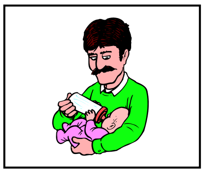


Figure 3.21: Items from Gap-fill task (Set A)

The participants were instructed to fill the gap with the correct article (***der*** or ***den***) and noun in German in order to make the sentence match the picture. There were four conditions; SVO sentences in which the subject (SVO+Subj; *k* = 6) or the object (SVO+Obj; *k* = 6) were missing, and OVS sentences in which the subject (OVS+Subj; *k* = 6) or object (OVS+Obj; *k* = 6) were missing. The participants were provided with each of the missing noun phrases written in English next to the gap. This re-emphasised that they not only needed to provide the missing noun but also the missing article. Before beginning the activity the participants were reminded that all of the missing nouns were masculine and they were, therefore, not allowed to use the words ***die*** or ***das*** for ***the***. It was anticipated that the participants would perform at ceiling level on the subject conditions (SVO+Subj, OVS+Subj) at pre-test, since the participants were already familiar with the nominative article ***der*** prior to the commencement of the study. Consequently any improvement in participant scores at post- and delayed-test would be attributable to an improvement in the learners' use of the new article ***den***.

As can been seen in Figure 3.21 each sentence was made up of one feminine or neuter noun phrase plus the missing masculine noun phrase. In the pilot version of this task sentences containing two masculine nouns were also used; however it was observed that for sentences containing only masculine nouns it would be possible for participants to correctly complete the activity by simply choosing the opposite article to the one already given in the sentence (e.g. ***den*** if the sentence contained ***der*** and vice versa). Such a strategy would result in the participants reaching the correct answer, but does not provide evidence that they are correctly processing and understanding the use of the respective articles. Therefore inclusion of feminine and neuter nouns prevented participants from employing such a strategy and provided a means of more rigorously testing whether participants were attending to the target feature when completing the task.

Scoring procedure

The total possible score on this task was 24. One point was awarded for each article correctly produced (one per test item). Six points were available for each condition (SVO+Subj, OVS+Subj, SVO+Obj, OVS+Obj). All of the test papers were marked by the researcher. As the focus of this task was on the learners' correct production of the article, participants were still awarded one point if they had supplied the correct article but had written the noun in English or had supplied the correct article but left out the noun (e.g. ***der*** dog).

3.6.4 Oral outcome measures

The oral tasks were interactive activities which were completed one-to-one with the researcher. The test stimuli for the three oral tasks were simple, transitive NVN constructions generated from a sub-list of nine masculine nouns (six animate and three inanimate) and six transitive verbs, which had been identified from the vocabulary list (Appendix 10). Soft toys (representing each of the nine nouns) were used within the tasks; either as part of the stimuli (*Act-Out Production* task, *Sentence Repetition* task), or as the method of response (*Act-Out Comprehension* task). Novel test items were generated from the sub-list of nouns and verbs for each version (A, B, C) of the three tasks within the split-block design. Therefore the participants would only have responded to a given test item once throughout the course of the study. The test items for the *Act-Out Comprehension* and *Sentence Repetition* task were pre-recorded by the researcher. The recordings were played to each participant via a laptop and headphones (a headphone splitter was used so that the researcher and participant could listen to the recording simultaneously). Each test item was played only once. The recording was paused by the researcher between each test item in order for the participant to respond. The participants' responses were recorded on a Dictaphone and a video camera in order to enable transcription at a later date.

3.6.4.1 *Act-Out Comprehension* task

The *Act-Out Comprehension* task constituted a test of the learners' aural comprehension ability. The task was modelled on the act-out tasks which have been used in previous studies to investigate young learners (e.g. 2 to 3 years old) cue reliance in their L1 (e.g. Chan, Meints, Lieven, & Tomasello, 2010; Dittmar, Abbot-Smith, Lieven, & Tomasello, 2008).

The *Act-Out Comprehension* task required the learners to listen to a transitive sentence in the target language and act out that sentence using the soft toys provided:

SVO sentence:

1) Participant hears — Der Bär umarmt den Tiger
   the-NOM bear hugs the-ACC tiger
2) Participant acts out — (e.g.) bear hugging tiger

OVS sentence:

1) Participant hears — Den Löwen kickt der Elefant.
   the-ACC lion kicks the-NOM elephant
2) Participant acts out — (e.g.) elephant kicking lion

Figure 3.22: Items from Act-Out Comprehension task (Set B)

In total the test contained 18 stimuli, which were counterbalanced for word order (SVO, *k* = 9; OVS, *k* = 9). As in the Sentence Matching task, this made it possible to determine whether the learners were relying solely on word order, as would be predicted by both the Competition Model (MacWhinney, 2001) and the FNP (VanPatten, 2004a), or the target grammatical feature when interpreting the target language input. As noted in section 3.6.2, the three animacy conditions were counterbalanced.

Scoring procedure

The action which had been performed by the participant (e.g. dog chasing table) was transcribed from the video recordings by the researcher and then scored. The total possible score on this task was 18. One point was awarded for each sentence acted out correctly (SVO = 9, OVS = 9). Given that this was a listening task which was performed under pressure (i.e. one-to-one with the researcher), allowance was made within the scoring for instances of the learners mishearing the stimuli or misunderstanding the lexical semantics of the verb. If the learner performed the sentence with the correct agent – patient relationship, but had incorrectly interpreted the verb, then one full mark was still given, for example:

1) Participant hears — Der Löwe verfolgt den Hund
   the-NOM lion chases the-ACC dog
2) Participant acts out — (e.g.) lion kicking dog

Similarly if the participant misinterpreted one of the nouns, but still assigned the correct agent / patient role to the remaining noun, then one full mark was still given, for example:

1) Participant hears — Der Tiger küsst den Affen.
   the-NOM tiger kisses the-ACC monkey
2) Participant acts out — (e.g.) tiger kissing elephant

Such instances were rare. If the participant misinterpreted two or more elements of the sentence (e.g. both nouns or verb + one noun) then a mark of zero was given.

3.6.4.2 *Act-Out Production* task

The *Act-Out Production* task was a test of the learners' ability to produce the target feature correctly when responding to visual stimuli, under time- and communicative pressure. J. White and Ranta (2002) observed that a majority of studies employing oral production tasks have tended to utilise tasks which involve production within "discrete-point, limited response items" (p. 267), for example naming objects within a picture. Such tasks could be described as "oral grammar tasks" (J. White & Ranta, 2002, p. 267) rather than tasks requiring production of language in order to convey a meaningful message (Lightbown, 2000). In the present study, the task was sentence-, rather than word-, level and soft toys were utilised as visual stimuli in order to keep the learners focussed on meaning rather than preoccupied with form. The format of the task was as follows:

1)	Researcher acts out	dog chasing ball
2)	Participant says	(e.g.) der-NOM Hund verfolgt den-ACC Ball
1)	Researcher acts out	lion kissing monkey
2)	Participant says	(e.g.) der-NOM Löwen küsst den-ACC Affen

Figure 3.23: Items from Act-Out Production task (Set B)

In total the participants produced 12 full sentences. For each sentence they were required to produce both the subject (***der***) and object (***den***) articles as all of the nouns were masculine. As with the *Gap-fill task* it was expected that the learners would correctly produce the subject article ***der*** at pre-test and further would overuse the subject article (***der***) for the object noun phrase at pre-test. Consequently any improvement made by the learners at post-test would be due to an improvement in their production of the target feature, the object article ***den***.

Scoring procedure

The Dictaphone recordings of the learners' responses on this task were transcribed and then scored by the researcher. The total possible score was 24: 12 points for correct production of the subject article and 12 points for correct production of the object article within each sentence. Analysis also investigated the learners' overuse of the respective articles (section 4.4.2.4). At post- and delayed post-test there were a small number of instances of the learners correctly producing a sentence in OVS word order. As this was not a requirement of the task, no extra points were awarded; however the instances in which this occurred were tallied for reference (Appendix 11).

3.6.4.3 *Sentence Repetition* task

A second measure of the learners' oral production ability was included in order to test whether the learners were able to correctly produce the target feature within an OVS sentence, since it was anticipated that the participants would primarily utilise SVO word order when producing sentences in the *Act-Out Production* task. The *Sentence Repetition* task was modelled on the Elicited Imitation tasks, which have been utilised in previous studies (e.g. Erlam, 2006, 2009; Harley & Hart, 2002; Myles & Mitchell, 2012). It is argued that when completing an Elicited Imitation task, learners will only be able to correctly repeat those grammatical features which are already part of their interlanguage (Erlam, 2006, 2009). As such it is thought to be a test of more implicit knowledge and the learners' underlying grammatical competence. The format of the *Sentence Repetition* task was as follows:

SVO sentence:

1) Researcher acts out — bear hugging elephant
2) Participant hears — Der Bär umarmt den Elefanten.
   the-NOM bear hugs the-ACC elephant
3) A two second beep immediately follows the sentence
4) Participant repeats the sentence

OVS sentence:

1) Researcher acts out — dog kicking tiger
2) Participant hears — Den Tiger kickt der Hund.
   the-ACC tiger kicks the-NOM dog
3) A two second beep immediately follows the sentence
4) Participant repeats the sentence

Figure 3.24: Items from Sentence Repetition task (Set C)

Six sentences were included (3 SVO, 3 OVS). Correctly producing the target feature within the OVS condition would indicate that the learner was not simply relying on a strategy such as ***der*** first, ***den*** second when producing target language sentences. As the sentences contained masculine nouns only, the learners were required to produce both the subject and object articles within all of the sentences, resulting in four conditions for scoring and analysis; SVO+Subj, OVS+Subj, SVO+Obj, OVS+Obj. With the change in word order the two articles swapped between sentence initial and medial position. Those features which are embedded within an utterance (i.e. sentence medial) are likely to be the most difficult to recall (Tomita, Suzuki, & Jessop, 2009). The sentences varied in length between six and nine syllables, in order to avoid the learners being able to memorise and subsequently correctly repeat the stimuli (Erlam, 2009). Additionally, an important feature of Elicited Imitation tasks is that the learner is initially focussed on the meaning of the utterance before repeating it (Erlam, 2006; Myles & Mitchell, 2012; Tomita et al., 2009); within the present study this was achieved by using the soft toys to act out the sentence, before each test item was heard. Additionally the participants had to wait for a short (2 second) beep before reproducing the sentence. This was in line with the recommendation that there is a short delay between the learner hearing and subsequently reproducing the stimuli (Erlam, 2006, 2009), in order to avoid a reliance on rote repetition. The inclusion of ungrammatical test items is also recommended (Erlam, 2006, 2009; Tomita et al., 2009), since the spontaneous correction of ungrammatical sentences serves as an additional indication that the target grammatical form has been integrated into the learners' interlanguage and further that the learners are reconstructing, rather than simply repeating the target sentences. Due to time constraints, however, it was not possible to lengthen the *Sentence Repetition* task. Nevertheless the pre-test scores revealed that the participants were performing at chance-level on this task (section 4.4.3), therefore suggesting that they were not completing the task solely via a reliance on rote repetition.

Scoring procedure

The learners' responses were transcribed from the Dictaphone recordings and scored by the researcher. Learners were awarded one point for each article correctly repeated, giving a maximum score of 12, three per condition (SVO+Subj, OVS+Subj, SVO+Obj, OVS+Obj). Given that the focus of the task was on whether the learners could correctly produce the target grammatical feature (***den***) and its nominative counterpart (***der***) and due to the fact that all of the nouns included in the task were masculine, any use of feminine or neuter articles i.e. ***die*** or ***das*** (as well as other articles e.g. ***the***, ***un***) was marked as incorrect and given zero points. The learners' article use across the four test conditions was also analysed to determine whether any overgeneralisation of ***den*** had occurred (section 4.4.3.4).

3.6.5 *Sentence Reconstruction* task

In line with previous studies (e.g. Alderson et al., 1997; Bouffard & Sarkar, 2008; Roehr, 2008) the *Sentence Reconstruction* task was designed to test the learners' ability to make use of and talk about grammatical rules and the grammatical role of particular items in the sentence, i.e. their metalinguistic knowledge. The participant was shown a picture and five words and was asked to correctly order the words to create a sentence (NVN construction) to match the picture (Figure 3.25). In order to avoid giving any clue as to the correct order of the words and more specifically the correct position of the articles, no punctuation was included and the capital letter[42] was removed from the start of the sentence:

[42] Capital letters were included on the nouns as it is a requirement in written German

der | **Mann** | **schreibt** | **den** | **Brief**

Figure 3.25: Item from Sentence Reconstruction task (Set A)

Once the participant had arranged the words into a sentence the researcher asked them to explain why they had chosen to place the words in that order. Particular focus was given to why the participant had put the articles in the chosen positions.

The task contained three sentences to be rearranged and discussed in three gender conditions:

a) masculine subject + masculine object
b) masculine subject + feminine / neuter object
c) feminine / neuter subject + masculine object

Condition a) tested the learners' verbalisable knowledge of the target feature (***den***) and its nominative counterpart (***der***). Conditions b) and c) provided the opportunity for the learners to extrapolate their knowledge of the target feature to sentences containing a feminine or neuter article (neither of which differentiate between the nominative and accusative cases; see section 2.5).

Scoring procedure

The participants' responses were transcribed and scored by the researcher from the Dictaphone and video camera recordings. Each learner received two scores for this task. Firstly the learners were marked on their ordering of the five words. A total of 6 points were available for this section of the task (two points per sentence). The full two points were awarded if the participant was able to correctly order the words on the first attempt without prompting from the researcher. If the participant initially ordered the words incorrectly, but then corrected their mistake of their own volition, two points were still awarded. One point was awarded if the participant was only able to order the words correctly following prompting from the researcher. Zero points were awarded if the chosen order of words was incorrect.

The second set of scores related to the explanations the learners gave for their chosen word order. Scoring was restricted to the learners' explanations relating to the position of the subject and object articles within the sentence. The learners' explanations were coded using a data-driven approach. The coding categories were identified as they emerged from the data (L. Cohen et al., 2011; Mackey & Gass, 2005), although it should be acknowledged that a number of themes were expected to surface such as 'gender', 'subject' or 'doing', 'object' or 'receiving (the action)' (Appendix 12). Once the learners' explanations had been coded, they were scored by the researcher. The total possible score was six points, with a maximum of two points awarded per sentence. One point was awarded for correct explanation of the function (i.e. assigning a thematic role) of each article within the sentence (e.g. ***der*** *goes with Hund because the dog is doing the chasing,* ***den*** *goes with Mann because he is being chased*). The use of metalinguistic terms (e.g. subject / object) was not required in order for an explanation to be marked as correct. If the learner was able to provide the correct explanation following prompting from the researcher, half a point was awarded. If the learner was unable to explain the function of the articles within the sentence then zero points were awarded. Explanations relating to gender (e.g. ***der*** *goes with Hund because it's male*), although grammatically correct, were not awarded any points since the focus was on ascertaining the learners' understanding of the target grammatical rule (i.e. ***der*** indicates subject, ***den*** indicates object).

Interrater reliability

Due to the fact that the numerical scoring of the learners' explanations was based on the researchers' coding of the qualitative data, it was deemed important to establish the reliability of the coding / scoring protocol (Mackey & Gass, 2005). Therefore a second marker was recruited to score a subset of the data. The subset consisted of 20 transcripts chosen at random[43] from the Year 6 post-test dataset (five transcripts from each of the Year 6 experimental groups in Schools 1 and 2).

Information about the activity in question and instructions on the scoring system to be used were provided for the second marker to read through at their own pace (Appendix 13). The second marker and researcher then scored one practice transcript[44] together, in order to check the second marker had understood the scoring system. The second marker then marked the 20 sample transcripts independently. All of the transcripts were anonymised and contained no identifying information about any of the participants (e.g. gender, school, experimental group). Once all of the transcripts had been marked, the researcher and second marker went through each transcript together in order to see if / where any disagreement arose. In addition

[43] The participants within each experimental group from the two Year 6 classes were ordered using the *randomise* - RAND() – and SORT functions in Microsoft Excel. The first five participants within each group were chosen.
[44] The practice transcript was chosen from the remaining pool of transcripts after the subset had been chosen

an Inter-rater agreement score was calculated using the following equation (Cohen et al., 2011, 201):

$$\frac{\text{Number of actual agreements}}{\text{Number of possible agreements}} \times 100$$

The number of actual agreements was 114 out of a possible 120[45]. This gave an inter-rater agreement score of 95%, which was substantially higher than both the suggested minimum percentage of agreement which can be considered good (75%) and ideal (90%) (Mackey & Gass, 2005, p. 244). The researcher and second marker discussed and reached agreement on the small percentage (5%) of disagreements which arose.

3.6.6 Vocabulary test

In line with previous studies (e.g. Jackson, 2007; Marsden & Chen, 2011) a short vocabulary task was developed for use at pre- and post-test in order to ascertain the learners' knowledge of the core vocabulary utilised within the experimental activities and ensure that the three experimental groups had an equivalent level of vocabulary knowledge prior to and following the intervention. Language learners are thought to have limited attentional and processing resources (N. Ellis, 2006; Svalberg, 2012; VanPatten, 1996, 2004a, 2007), therefore if available processing resources are expended primarily on vocabulary recognition and comprehension, then it is unlikely that new grammatical forms within the input will be processed (see Availability of resources principle; VanPatten, 2004a). Therefore prior to implementing the interventions it was important to determine whether the learners across the TE-FMC, TE-F and Control groups had an adequate and similar level of vocabulary knowledge.

The vocabulary task was a test of receptive vocabulary knowledge. The learners were presented with a list of German nouns (k = 14) and verbs (k = 6) and four multiple choice options:

1.	**Vater**	a. father	b. grandfather	c. son	d. mother
2.	**Schwein**	a. horse	b. sheep	c. pig	d. cow
	…				
15.	**verfolgt**	a. chases	b. greets	c. loves	d. scares
16.	**füttert**	a. eats	b. feeds	c. telephones	d. visits

Each class worked through the vocabulary task altogether. The researcher read aloud the German word and the learners were instructed to circle their answer. In this way the target word was provided both aurally and in writing. It was deemed beneficial to utilise both modalities as the intervention activities and outcome measures included both listening and reading tasks. It is important to note however that the learners were working individually,

[45] Number of possible agreements: 120 = 20 pupils x 3 questions x 2 points per question

noting their responses on individual worksheets and no discussion was permitted during the test. Two versions of the vocabulary test were created and rotated between the experimental and year groups at pre- and post-test (Table 3.14). In total, the two versions of the vocabulary task included 40 key lexical items utilised within the experimental activities (28 nouns, 12 verbs). The two versions were created by assigning equivalent pairs of terms to opposite versions (e.g. *Vater* to task A, *Mutter* to task B), or by assigning words of equivalent transparency (i.e. similarity to English word; judged by the researcher) to opposite versions of the task (e.g. low transparency: *Kaninchen* to task A, *Schmetterling* to task B).

Table 3.14: Rotation of vocabulary test (A and B)

Group	*School*	*Age*	*Pre-test*	*Post-test*
TE-FMC	**1**	**Year 6**	A	B
		Year 5	B	A
	2	**Year 6**	B	A
		Year 5	A	B
TE-F	**1**	**Year 6**	A	B
		Year 5	B	A
	2	**Year 6**	B	A
		Year 5	A	B
Control	**3**	**Year 6**	A	B
	2	**Year 5**	B	A

Once the vocabulary test had been completed, the learners swapped their worksheet with their neighbour and scored one another's responses. Each class went through the test items together led by the researcher and individuals were chosen to provide the correct answer. This format meant that any learners who were unsure of terms were reminded of the correct meaning prior to beginning the experimental activities. Each correct answer was awarded one point, giving a maximum score of 20. The test sheets were collected by the researcher and the scores collated.

3.6.7 Grammatical sensitivity test

Grammatical sensitivity (see section 2.4) was measured using a sub-test (*Matching Words*) within the Modern Language Aptitude Test – Elementary (MLAT-E) (Carroll & Sapon, 2002). The MLAT-E is an adaptation of the original MLAT (Carroll & Sapon, 1959), which is appropriate for use with children aged 8 to 12 years old and is designed to assess an individual's "probable degree of success in learning a foreign language" (Carroll & Sapon, 2002, p. 2). Notably, criticisms have been raised of the MLAT, and its derivatives (e.g. MLAT-E), for example due to the fact that it does not include a test of working memory (e.g. reading span test) (Sawyer & Ranta, 2001), or a test of learners' inductive language learning ability (Skehan, 2002), despite it being one of the four main components of Carroll's (1981) construct of aptitude. Further, a written test such as the MLAT is heavily reliant on an individual's L1 literacy level (Milton & Alexiou, 2006; Tellier & Roehr-Brackin, 2013a). Based on such criticisms, alternative aptitude tests have been designed, such as the Pimsleur Language Aptitude Battery (PLAB) (Pimsleur, 1966), which includes measures of inductive language learning ability and motivation / interest in language learning, or Milton and Alexiou's (2006) test of more general cognitive ability, which utilises "picture-based and game-like activities" (p. 185) and consequently precludes a reliance on literacy skills. Notably, however, PLAB is recommended for use with older learners (aged 12 to 18) than those in the present study, and Milton and Alexiou's tests are designed for use with younger learners aged 5 to 7 years old.

For the purposes of the present study, the *Matching words* sub-test within the MLAT-E (British English version) was deemed the most appropriate measure to elicit information regarding the learners' grammatical sensitivity. This decision was in line with the precedent set in previous studies which have investigated the relationship between grammatical sensitivity (as measured by the MLAT) and instructed language learning (e.g. Robinson, 1997; Tellier & Roehr-Brackin, 2013a; VanPatten & Borst, 2012). The *Matching words* test "is designed to measure sensitivity to grammatical structure … and teach the examinee to recognise the job that a particular word does in a sentence" (Carroll & Sapon, 2002, p. 2). Such a skill is in line with the focus of the instruction within the present study, namely teaching reliance on case marking in order to assign grammatical roles within a sentence.

Format of the task

The MLAT-E manual (including instructional CD) and test papers were purchased from Second Language Testing, Inc. in May 2011 and April 2013 respectively.

The *Matching words* test (as well as the other sub-tests within the MLAT-E) are administered via the instructional CD. Information is provided regarding the format of the task and question examples are presented and explained. The participants work through a set of practice questions with the CD, which are designed to give practice in identifying the four target grammatical functions or "jobs" (subject, object, verb, adjective). Feedback is provided via the CD following each set of practice questions. The participants are then instructed to complete the test in the booklet provided.

In the *Matching words* test itself the participant is presented with pairs of sentences. In the first sentence a keyword is shown in capital letters. The participant is asked to identify the word within the second sentence which "does the same job", i.e. performs the same grammatical function, as the key word:

A small BOY rang the bell.

Our dog never bites the postman

☐ ☐ ☐ ☐ ☐ ☐

Figure 3.26: Item from Matching words task

In total the test contains 30 sentence pairs, which utilise a range of syntactic structures. Notably the *Matching words* task does not involve the use of any grammatical terminology (Carroll & Sapon, 2002), rather the learners are encouraged to make what Carroll (1981) describes as 'grammatical analogies' (p. 105) by identifying words which perform a similar function. The participants are able to work through the test at their own pace. The CD allows a total of 18 minutes in which to complete the test. All of the participants completed the test within this time frame.

Scoring procedure

The participants' responses on the *Matching words* task were scored by the researcher using the scoring sheets provided with the MLAT-E Manual. One point was available per sentence pair, giving a maximum of 30 points.

3.6.8 Administering the outcome measures

i) Written tasks

The two written tasks, the vocabulary test and the grammatical sensitivity task were administered by the researcher to each whole class during the regular German lesson. The two written tasks took approximately 20 minutes to complete, the vocabulary test approximately 10 minutes, and the grammatical sensitivity task 28 minutes (10 minutes instruction plus 18 minutes for test completion).

ii) Oral and metalinguistic tasks

The oral and metalinguistic tasks were completed one-to-one with the researcher and took approximately 15 to 20 minutes. Where possible, the oral tasks were administered before the participants had completed the written activities. This was due to the nature of the tasks, since the oral tasks were designed to elicit more implicit knowledge whereas the written tasks were more explicit in nature. However it should be noted that due to practical difficulties (i.e. school timetables, timing of German lesson) this was not always possible and a small number of participants at each time point were required to complete the oral activities after the written tasks. In such cases, care was taken to ensure that the oral tasks were not completed on the same day as the written tasks, rather at least one day later.

Due to the number of participants and the time needed to complete the oral activities with each participant individually, three research assistants were recruited to assist the researcher in administering the oral tasks. The research assistants were two MA students from the university and one lecturer, all from the Education department. Prior to conducting the oral activities in school, the researcher met with the research assistants to talk through the protocol for the activities. At the end of the meeting the research assistants were asked to demonstrate each of the activities (with the researcher acting as participant) in order to ensure that they were confident in the requirements of each task. In addition a comprehensive written protocol (Appendix 14) was provided for both the researcher and research assistant to follow when conducting the oral activities, to ensure that the tasks were delivered systematically and reliably. On any one day, one research assistant would be in school conducting the oral tasks alongside the researcher. Participants were assigned to either the researcher or research assistant using the following method: firstly the names of the participants in each class, for each experimental group separately, were randomly ordered. Name 1 was then assigned to the researcher, name 2 to the research assistant, name 3 to the researcher and so on. This method resulted in a random cross section of each class, counterbalanced across conditions, completing the oral activities with the researcher or research assistant at any one time point.

3.6.9 Missing data

When administering the intervention and outcome measures, one issue which arose was that of attrition. There were several occasions during the course of the study when one or more students were unable to attend a lesson due to absence from school, or other school commitments. Every effort was made to complete the activities with those learners on a separate occasion; nevertheless this was not always possible. Consequently 11 participants were excluded completely from the final data pool due to one, or a combination of, the following reasons:

- Participant had missed 2 or more intervention sessions
- Participant was not present for the post- and / or delayed post-test
- Participant had missed a significant amount of pre-teaching

In addition the data from one Year 5 participant from school 1 was excluded, since English was not this participant's L1.

When administering the outcome measures, there were a number of participants who were not present during the lesson in which the written tasks were completed, but who were able to complete the oral activities and vice versa. This resulted in differing numbers of participants having completed each task. Consequently the participants were included in the analysis of the individual written tasks, if they had completed both written tasks at post- and delayed post-test, regardless of whether or not they had also completed all of the oral tasks and likewise for the oral tasks. For the analysis in which the participants' performance across the outcome measures was compared, only those learners who had completed all of the outcome measures at both post- and delayed post-test were included:

Table 3.15: Number of participants included in analysis of tasks

	Written	Oral	All
TE-FMC	45	45	44
TE-F	38	41	38
Control	50	46	-
Total	133	132	82

Those participants who had not participated in the pre-test, but had been present for both the post- and delayed post-test, were included in the analysis. The mean pre-test score achieved by the participant's class was calculated for each activity and substituted as the pre-test score for the individual in question.

3.7 Statistical analysis

3.7.1 Parametric versus Non-parametric

Before analysing a set of data it is important to decide whether parametric or non-parametric tests will be used. The use of parametric tests in data analysis requires that certain underlying assumptions are met:

1) data is normally distributed
2) group variances are equal
3) data are interval
4) data are independent

The assumptions of normal distribution and homogeneity of variance will be dealt with in sections 3.7.1.1 and 3.7.1.2 respectively. With regards to assumptions 3 and 4, in the context of the present experiment, the assumption of independence would require that the data from each participant was not influenced by another participant. In this study each participant was working independently when completing the activities at pre-, post-, and delayed post-test, therefore this assumption was met. Further, all of the data used in the analysis was scale (or continuous), therefore the assumption that data is measured at least at the interval level was also met.

The question of whether a given data set satisfies the above assumptions is often overlooked in second language research (Larson-Hall, 2010, p. 74). Although it has been argued that parametric tests are robust enough to withstand violations of these assumptions

(Larson-Hall, 2010), others have maintained that the use of parametric tests when the underlying assumptions are not met could increase the possibility of making a Type II error (i.e. finding no relationship between variables when it does in fact exist) (Field, 2009; Larson-Hall, 2010). For instances in which the assumptions of parametric tests are not met, an alternative approach would be to use the non-parametric counterpart. Non parametric tests, sometimes referred to as 'assumption-free tests' (Field, 2009, p. 540) or distribution-fee tests (Howell, 2010, p. 660), carry fewer assumptions than parametric tests. Rather than relying on the data being normally distributed, they function through creating a rank-order for the points within a dataset and subsequently analysing these ranks rather than the data itself (Field, 2009; Howell, 2010). Further, non-parametric tests are insensitive to the influence of outliers as they make use of the median rather than the mean (Field, 2009; Howell, 2010; Larson-Hall, 2010). Consequently, although non-parametric tests are sometimes thought to have lower power relative to the equivalent parametric test, for datasets which do not meet the necessary parametric assumptions they may in fact have greater power to find a statistical difference (Howell, 2010; Larson-Hall, 2010).

3.7.1.1 Normality of distribution

The assumption of normality assumes that all data points for a given test are distributed evenly around the centre of all the scores (i.e. measure of central tendency). When presented graphically the data is typically characterised by a 'bell-shaped' curve (Field, 2009). It has been argued that utilising datasets of larger than 30 (as the present study does) can obviate the need to test for normality (Weinberg & Abramowitz, 2002, p. 276). Nevertheless evidence has also been put forward that even with datasets of more than 160 parametric tests may not be sufficiently robust when the distribution is skewed (Wilcox, 1998). This assumption was therefore tested before proceeding with any analysis.

Determining whether a given dataset is normally distributed can be achieved in a number of ways. Within the present study a graphical representation of the data (histogram) was generated in order to provide a visual indication of the normality of distribution (Larson-Hall, 2010, p. 74). In addition, a normality test was also run in order to provide a more robust measure of the nature of the distribution for each set of data. There are two possible normality tests: the Kolmogorov-Smirnov test and the Shapiro-Wilk test. Both tests determine the normality of a given dataset by calculating whether the sample data significantly deviates from an equivalent normally distributed set of data with the same mean and standard deviation (Field, 2009). A significant finding ($p < .05$) would suggest that the dataset does deviate from normality. It is important to note that normality tests are not necessarily 100% conclusive in determining whether a non-normal distribution will bias any statistical analysis used, since with a large sample size even minimal deviation from normality can result in a significant finding (Field, 2009). Therefore Field (2009) and Larson-Hall (2010) advocate the use of both graphical and numerical measures when assessing the distribution of a given dataset.

With regards to the present study the results of the Shapiro-Wilk test were utilised as this test is thought to be more accurate[46] (Field, 2009, p. 546) and has been shown to have more power than the Kolmogorov-Smirnov test (Razali & Wah, 2011).

3.7.1.2 Homogeneity of variance

A second critical assumption is that of homogeneity, or equality, of variance (Field, 2009; Larson-Hall, 2010). Homogeneity of variance refers to the assumption that when there are two or more groups present within a dataset, the data from each sample group will have the same amount of variance (Larson-Hall, 2010). Variance can be defined as the distance from the mean to any point and is a measure of the 'dispersion' of data points around the mean (Larson-Hall, 2010, p. 87). The homogeneity of variance within a dataset can be examined numerically using Levene's test of homogeneity of variance[47], which functions by testing the null hypothesis that variances within a given dataset are equal. If the test result is significant ($p < 0.05$), then the null hypothesis should be rejected and it should be assumed that the variances are not equal (Larson-Hall, 2010). Within the present study the homogeneity of variance within the datasets was measured using Levene's test.

3.7.2 Non-parametric analysis

3.7.2.1 Non-parametric tests

All of the data collected within the present study was found to violate the assumption of normality and a majority violated the assumption of homogeneity of variance (section 4.1). Consequently non-parametric tests were utilised to analyse the performance of the experimental groups over the course of the study (RQ1) and to compare between groups (RQ3) (Table 3.16, over page). The parametric Factorial Mixed ANOVA would ordinarily be employed to carry out analysis involving both Within-group (Time) and Between-group (Group) factors, enabling a simultaneous comparison of both the improvement of each group over the time points and of the difference between each groups' performance at each time point (Field, 2009; Larson-Hall, 2010). However, there is no non-parametric equivalent of this test; therefore, individual non-parametric tests were employed to carry out each stage of the analysis. Nevertheless, in order to maintain parity with previous research (which has primarily utilised parametric, rather than non-parametric analysis) the equivalent parametric tests were

[46] The Shapiro-Wilk test gives an exact significance value, whereas the Kolmogorov-Sminov test sometimes gives an approximate significance of $p = .2$ (Field, 2009, p. 546).
[47] Homogeneity of variance can also be examined graphically through the use of boxplots.

also run and the results were found to support those yielded from the non-parametric analysis. Only the results of the non-parametric tests are reported. The results of the parametric analysis are available if required, but are beyond the scope of this thesis.

Table 3.16: Non-parametric tests utilised in the study

Test	Between- or Within-group	Levels	Post-hoc	Purpose
Kruskall Wallis	Between-group	3	Pairwise comparisons / bonferroni correction^{+}	Compare TE-FMC, TE-F and Control groups' performance at each time point
Mann Whitney U	Between-group	2	N/A	Compare TE-FMC and TE-F groups' performance at each time point
Wilcoxon signed rank	Within-group	2	N/A	Analyse each group's (TE-FMC, TE-F, Control) performance over pre- and post-test
Friedman's ANOVA	Within-group	3	Pairwise comparisons / bonferroni correction^{+}	Analyse TE-FMC and TE-F groups' performance over pre-, post- and delayed post-test

$^{+}$see section 3.7.2.2

The Control group did not participate in the delayed post-test; therefore two waves of analysis were conducted. Firstly the pre- and post-test performance of all three experimental groups (TE-FMC, TE-F, Control) was analysed in order to establish whether a) the TE-FMC and TE-F groups improved following the intervention and b) the non-active Control group made any gains at post-test, thereby eliminating a potential test effect. Secondly the TE-FMC and TE-F groups' performances were analysed and compared over the three time points (pre-, post-, delayed post-test) in order to determine whether any observed improvement was sustained nine weeks after the respective intervention.

3.7.2.2 Statistical significance

Inferential statistics, such as the non-parametric tests detailed above, provide a means of testing whether the effect of the independent variable (e.g. experimental group) on the dependent variable (e.g. performance on an outcome measure) is due to chance. As such inferential statistical tests generate a *p*-value, or probability value. The *p*-value tells you "how likely it would be that you would get the difference you did (or one more extreme), by chance alone, if there really is no difference between the categories presented by your groups" (Robson, 2011, p. 446). In a majority of SLA research the *p*-value tends to be set (arbitrarily) at 5%, meaning that there is 95% certainty that a given result is not due to chance (Field, 2009, p. 50). As such the alpha level (cut-off point for *p*-value) is .05. If a given statistical test generates a *p*-value which is less than the alpha level ($p < .05$) then that finding is said to be significant; there is a statistically significant difference between the two groups compared and less than a

5% chance that this finding is due to chance. In line with previous research, the alpha level was set at $p < .05$ for the analysis conducted within the present study. It is important to note that when post-hoc pairwise comparisons were conducted (Table 3.16), the alpha level was adjusted by applying a Bonferroni correction. The standard alpha level .05 was divided by the number of comparisons made (e.g. over time: three comparisons resulting in an adjusted alpha level of $p < .0167$). Applying such a correction controls the familywise error rate for multiple comparisons of a single dataset and retains an overall Type I[48] error rate across all of the comparisons of .05.

It is important to consider, however, that whilst the *p*-value provides an indication of whether a given finding was due to chance, it does not speak to the size of an observed effect or relationship (Norris & Ortega, 2000) or whether a given effect is meaningful within the research context (Field, 2009; Larson-Hall, 2010). Equally a non-significant finding ($p > .05$) does not necessarily indicate that no effect was present within the data, although it is often interpreted as such (Norris & Ortega, 2000). It is therefore important to report the accompanying descriptive and inferential data, so that the results of significance tests can be interpreted accurately (Field, 2009; Larson-Hall, 2010; Norris & Ortega, 2000; Robson, 2011).

Norris and Ortega (2000) emphasise that "the fundamental problem with the use of statistical significance tests in L2 type-of-instruction research is that such tests are not designed to provide answers to the primary research questions of the domain" (p. 494), for example how effective a given treatment was, or the effectiveness of one type of instruction compared to another. In order to answer such questions additional information as to the magnitude of an observed statistically significant effect are needed. Given that such questions are the focus of the present study. Calculations of effect size were incorporated into the analysis conducted and are detailed in the next section.

3.7.2.3 Calculating effect size

Utilising a standardized measure of effect size such as Cohen's *d* can give a meaningful indication of how large the difference between two means is and enables comparison of the effectiveness of different instructional treatments across multiple studies (Howell, 2008; Norris & Ortega, 2000). Therefore the reporting of effect sizes is strongly encouraged (Field, 2009; Howell, 2008; Larson-Hall, 2010; Norris & Ortega, 2000) and was included in the present thesis.

Cohen's *d* can be calculated using basic descriptive statistics: the dependent variable means, standard deviations of the contrasted groups, and group sample sizes. For the present

[48] Type I error refers to incorrectly finding a relationship between variables which does not exist (i.e. incorrectly rejecting the null hypothesis)

study Cohen's *d* was calculated using the equation set forth in Norris and Ortega's (2000) meta-analysis:

$$d = \frac{mean_e - mean_c}{S_w}$$

$mean_e$ refers to the experimental group mean score, $mean_c$ refers to the comparison group mean score, and S_w refers to the pooled standard deviation of the two groups. The use of the pooled standard deviation is recommended as an alternative to using the Control group standard deviation only, as the standard deviation of any one particular group could be susceptible to sampling error (Norris & Ortega, 2000). The pooled standard deviation (S_w) can be calculated using the following equation (Norris & Ortega, 2000), in which *n* refers to the sample size of either group and *S* the standard deviation of either group:

$$S_w = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 - 1) + (n_2 - 1)}$$

Cohen's *d* was calculated using the above equations for each of the outcome measures included in the present study, in order to determine:

a) the magnitude of the observed difference between the TE-FMC and TE-F groups and the Control group
b) the change in each group's scores between the pre-test and the post- and delayed post-test

Cohen's *d* was considered to construe a small effect if $0.2 < d < 0.5$, a medium effect if $0.5 < d < 0.8$ and a large effect if $0.8 < d$.

Additionally a mean effect size was calculated across all of the dependent variables for each group in order to estimate the overall average effect attributable to each instructional treatment. In order to establish the statistical 'trustworthiness' of mean effect sizes, it is recommend that 95% confidence intervals are calculated (Norris & Ortega, 2000, p. 449). Therefore the following equation was utilised to this end[49] (Norris & Ortega, 2000, p. 505):

$$\mathrm{CI} = d \pm 2 \frac{\mathrm{SD}}{\sqrt{\mathrm{k}}}$$

3.7.3 Principal component analysis

In line with previous studies (R. Ellis, 2005; Han & Ellis, 1998; Marsden & Chen, 2011) the present study employed Principal Component Analysis (PCA) in order to explore the pattern

[49] *k* refers to the number of dependent variables contributing effect sizes

underlying the participants' performance across the six outcome measures and to investigate the predictions made regarding the knowledge types likely to be elicited from each of the tasks (RQ2). The aim of PCA is to determine whether a group of variables are driven by the same underlying, sometimes referred to as *latent*, variable (Field, 2009) and does so by analysing the size of the correlation between each of the variables. The presence of large correlation coefficients between subsets of the variables would suggest that "those variables are measuring the same underlying construct", or *factor* (Field, 2009, p. 628).

3.7.3.1 Determining the suitability of data for PCA

Before running PCA it is important to determine whether the given dataset is appropriate for analysis. Suitability is assessed based on two components, sample size and the strength of the correlations between variables. Firstly correlation coefficients can fluctuate particularly with small sample sizes, therefore utilising a sufficiently large sample is crucial in PCA (Field, 2009). It is commonly suggested that at least 10 to 15 participants are required per variable, although the empirical basis for this proposal is unclear (Field, 2009). Tabachnick and Fidell (2007) proposed that a minimum of 300 cases[50] is acceptable and will result in a stable factor solution. Therefore this was used as a guideline figure in the present study. Further the Kaiser-Meyer-Olkin measure of sampling adequacy can be employed. This test yields a KMO value of between 0 and 1; a value of 0.5 is barely acceptable (Kaiser, 1974, cited in Field, 2009), whereas a value between 0.5 and 0.7 is mediocre, between 0.7 and 0.8 is good, and greater than 0.8 is considered very good (Hutcheson & Sofroniou, 1999, cited in Field, 2009).

Secondly, since the aim is to explore the underlying relationship between groups of variables, PCA requires that there are sufficiently strong correlations between the variables in question (although if the correlation between variables is too strong or too weak this can also be problematic). The correlations can be examined in two ways; firstly the *R*-matrix (matrix of correlations), which is generated when the PCA is run, can be visually examined, along with the *R*-matrix determinant (*R*). An *R* greater than .00001 indicates that there is not severe multicollinearity (i.e. very highly correlated variables) within the dataset. Secondly Bartlett's test of sphericity can be run to test the null hypothesis that the variables do not, or only very weakly, correlate (Field, 2009). A significant result ($p < .05$) would indicate that the null hypothesis can be rejected, and that there is a sufficient level of correlation between the variables in question.

All of the methods detailed above were utilised in the present study to assess the suitability of the datasets for analysis via PCA.

[50] In the present study, total cases = number of participants x number of variables (tasks)

3.7.3.2 Extracting factors

One important decision to be made when interpreting PCA output relates to the number of factors (or components) to be retained. For each factor extracted, an eigenvalue is generated which indicates the importance of that factor within the factor solution. This is based on the amount of variance within the dataset which can be explained by that factor (Field, 2009). Kaiser's criterion states that extracted components with an eigenvalue greater than 1 should be included within a factor solution, as they are likely to represent a substantial amount of variation (Field, 2009). Alternative criterion have been proposed: Jolliffe (1972, cited in Field, 2009, p. 640) for example suggested that all factors with an eigenvalue greater than 0.7 should be retained. Field (2009), however, proposes that Kaiser's criterion may be more accurate when fewer than 30 variables are included in the analysis. Kaiser's criterion was therefore adopted for the purposes of the present study.

Factor loadings are calculated for each variable within the dataset. The factor loading is the correlation between a variable and a given factor and further indicates how important the variable is to the factor in question (Field, 2009). Factor loadings range between 0 and 1; Stevens (2002) recommends that a factor loading larger than 0.722 can be considered significant for a sample size of 50, whereas for a sample of 100, 0.512 can be considered significant[51].

Once an initial factor solution has been extracted from the analysis, a *factor rotation* can be applied in order to aid interpretation of the output. In an unrotated solution, a majority of the variables are likely to load heavily onto one of the extracted factors, but also exhibit small loadings onto the other factors. A *factor rotation* essentially rotates the factor solution "such that variables are loaded maximally to only one factor" (Field, 2009, p. 642). Two types of *factor rotation* can be employed, orthogonal or oblique. The main difference is that in oblique rotation the extracted factors are permitted to correlate, therefore oblique rotation is the preferred method if there is a theoretical reason why the underlying factors could be related (Field, 2009, p. 643). In the present study the variables analysed within the PCA were the outcome measures, which had been developed as tests of either more explicit or more implicit knowledge. Notably however it is likely that a test will tap into both types of knowledge to a certain extent (R. Ellis, 2002, 2005, 2009b; Roehr, 2008). Further the constructs of implicit and explicit knowledge are likely to be correlated to a certain extent (Isemonger, 2007). Therefore oblique rotation was deemed the most appropriate method for use with the present study dataset. In addition the direct oblimin, rather than the promax, approach to

[51] Significance of factor loadings is based on an alpha level of $p < .01$ (Field, 2009; Stevens, 2002)

oblique rotation was utilised, since the promax approach is only recommended for use with very large datasets.

3.7.4 Exploring individual participant performance on the outcome measures: Chi-square and McNemar test

As well as analysing each group's performance on the outcome measures, more fine-grained analysis was conducted in order to investigate the nature of the deviation within each group's scores at post- and delayed post-test. Within the TE-FMC and TE-F groups the learners were divided into sub-groups, based on their scores on each task: Got It, Middle, or Not Got It (see section 5.2). Analysis was conducted in order to explore the distribution of learners across the sub-groups using the following tests:

Table 3.17: Tests utilised to analyse categorical sub-group variable

Test	Between- or Within-group	Levels	Purpose
Chi-Square	Between-group	2 x 2[+]	Compare number of Got It versus Not Got It learners in TE-FMC and TE-F groups
McNemar	Within-group	3 x 2[52]	Compare number of learners in Got It, Middle, Not Got It sub-groups at post- and delayed post-test (for TE-FMC and TE-F groups separately)

[+]Middle sub-group was not included in this analysis

Both the Chi-Square and McNemar tests constitute non-parametric tests which are suitable for use with categorical (or nominal), rather than continuous data, i.e. data in the form of counts (Larson-Hall, 2010).

The Chi-Square test compares "the frequencies you observe in certain categories to the frequencies you might expect to get in those categories by chance" (Field, 2009, p. 688), i.e. the observed count is compared to the expected count for a given categorical variable. In the current study this test was utilised in order to investigate whether at post- and delayed post-test respectively, there was a difference in the observed count of learners in the Got It and Not Got It sub-groups as a function of experimental group (TE-FMC, TE-F). Use of Chi-Square requires that two assumptions are met: firstly the expected count for a given variable should be greater than 5 (Field, 2009). Secondly the data within each cell of the contingency table[53] are independent from one another (i.e. one participant can only contribute to one cell) (Field, 2009; Larson-Hall, 2010), therefore it is not appropriate for use within a repeated

[52] This analysis has since been redone for all tasks with the Middle sub-group excluded, since the McNemar test is for use with dichotomous variables. The same pattern of results was found throughout (see Hanan & Marsden, in progress).
[53] The contingency table presents the observed and expected counts for each categorical variable

measures design. Consequently the McNemar test was utilised in order to carry out the analysis, in which the distribution of learners across the three sub-groups was compared between post- and delayed post-test.

3.7.5 Exploring the relationship between the outcome measures and the grammatical sensitivity task

In line with previous studies (e.g. VanPatten & Borst, 2012), three statistical techniques were employed in order to explore the relationship between the TE-FMC and TE-F participants' performance on the grammatical sensitivity task and the individual outcome measures: ANCOVA, correlation, simple regression. These tests constitute parametric tests; however there are no non-parametric alternatives. Therefore, despite the non-normal distribution of the data, the three tests were utilised but the results are interpreted with caution.

3.7.5.1 ANCOVA

In an analysis of covariance (ANCOVA), an additional independent variable (covariate) is included in the analysis in order to reduce the amount of unexplained variance in the dataset (Field, 2009; Larson-Hall, 2010). As such the ANCOVA can be utilised as a means of exploring whether a covariate may have exerted an independent effect on the dependent variable (Larson-Hall, 2010). Alongside the assumptions of normality and homogeneity of variance (section 3.7.1), several assumptions must be met in order to run an ANCOVA. Firstly, the covariate can be a continuous or a categorical variable (Larson-Hall, 2010). Secondly, the assumption of homogeneity of regression slopes must be met, which refers to the requirement that the relationship between the covariate and the dependent variable must be similar between all of the groups included in the analysis (Field, 2009; Larson-Hall, 2010). For example, in the present study, if a positive relationship was observed between grammatical sensitivity task scores and scores on the *Sentence matching* task for the TE-FMC group, a positive relationship must also be observed for the TE-F group. To address this assumption, scatterplots can be generated in order to visually examine the linearity of the dataset. Further, the data can be tested for an interaction between the covariate and the independent or treatment variable. If a statistically significant ($p < .05$) interaction is found then an ANCOVA cannot be used. With regards to the present study, no interaction was observed between the learners' grammatical sensitivity scores and their group membership for any of the outcome measures, thereby satisfying the homogeneity of regression slopes assumption.

3.7.5.2 Correlation and standard regression

Whilst an ANCOVA can be utilised in order to investigate whether or not a given variable may have influenced the dependent variable, regression analysis provides a means of determining *the extent to which* one or more explanatory (or predictor) variable(s) may have influenced the participants' performance on the response (or dependent) variable. Regression analysis does so by calculating the model which best fits the dataset (i.e. line of best fit), and determining the amount (%) of variance which can be accounted for by the specified explanatory variable (Field, 2009). The "goodness of fit" of the model can be assessed based on the correlation coefficient (r) and the R^2 value generated for the model.

Firstly, Pearson's correlation coefficient (r) denotes the relationship between the explanatory variable and the dependent variable. The correlation coefficient can range from -1 to 1 and indicates both the strength of the relationship (a coefficient close to 1 equals a strong relationship) and the direction of the relationship (whether the relationship is positive or negative) (Field, 2009; Larson-Hall, 2010). Secondly the R^2 value represents "the amount of variance in the outcome explained by the model" (Field, 2009). The R^2 value generated will range from 0 to 1, and can be multiplied by 100 in order to ascertain the percentage of the variation within the dependent variable which can be accounted for by the explanatory variable.

In the current study, simple regression analysis was also employed in order to investigate how much of the variation within each group's scores on the respective outcome measures could be accounted for by the learners' performance on the test of grammatical sensitivity.

Chapter 4: Results 1

This chapter details analysis of the learners' performance on each of the outcome measures utilised in the study. Sections 4.1 and 4.2 outline the preparatory analysis carried out. First, a summary of the findings of the tests of normality and homogeneity is presented (section 4.1). Secondly, the equivalence of the learners' pre-test performance by age (Year 5, Year 6) and school (School 1, 2, 3) is established (section 4.2.1). Thirdly, the three groups' (TE-FMC, TE-F, Control) pre- and post-test performance on the vocabulary test is compared (section 4.2.2). The analysis of the individual outcome measures is then presented; section 4.3 provides the results of the two written tasks and section 4.4 the three oral tasks. The following analysis is presented for each individual outcome measure:

a) Analysis of learners' total scores
b) Analysis of learners' score by test condition (e.g. SVO, OVS)
c) Examination of effect sizes

It is also important to note that the Control group did not participate in the delayed post-test; therefore each set of analysis is presented in the following order:

i) Over time (pre- to post-test); including TE-FMC, TE-F and Control groups (Wilcoxon signed rank test)
ii) Over time (pre-, post- to delayed post-test); including TE-FMC and TE-F groups (Friedman's ANOVA test)
iii) Between groups (pre- and post-test); including TE-FMC, TE-F and Control groups (Kruskall Wallis test)
iv) Between groups (delayed post-test); including TE-FMC and TE-F groups (Mann Whitney U test)

Section 4.5 reports the results of the analysis by Animacy condition for the Act-Out Comprehension task[54]. Finally, the quantitative (section 4.6.1) and qualitative (section 4.6.2) analysis of the metalinguistic task is presented.

[54] Since the same pattern of results was found across all of the outcome measures, only the results of the Act-Out Comprehension task are presented (section 4.5). The descriptive statistics for the Sentence Matching, Gap-fill and Act-Out Production tasks can be found in Appendix 24.

4.1 Normality and homogeneity of the dataset

Prior to carrying out any analysis it was important to establish the normality and homogeneity of the dataset, in order to guide the decision as to whether parametric or non-parametric tests were to be used (see section 3.7.1). The analysis of normality and homogeneity was conducted on the datasets for the three experimental groups (TE-FMC, TE-F, Control) separately, since subsequent analysis compared the performance of these three groups on each task. In addition, given that analysis was also conducted in order to establish the equivalence of the two year groups' (Year 5, Year 6) performance as well as the three schools' (1, 2, 3) performance at pre-test, the normality and homogeneity of the Year 5 and Year 6 pre-test data was examined for each task, along with the data for Schools 1, 2 and 3.

Normality of distribution

The Shapiro-Wilk test of normality revealed that, for all tasks, the datasets (by experimental group, by age group, by school) were non-normally distributed (Appendix 15). Histograms were generated for each task and group[55] and were found to corroborate the findings of the Shapiro-Wilk test.

Homogeneity of variance

The results of Levene's test revealed that the variance within the pre-test data for the Sentence Repetition task and Sentence Reconstruction (Explanation) task, as well as the post-test data for all of the activities, was significantly different between the three experimental groups. Equal variances were observed between the TE-FMC, TE-F, and Control groups' data for the remaining tasks at pre-test and between the TE-FMC and TE-F groups' data for all of the delayed post-test datasets. When analysed by age group, Levene's test was non-significant for all of the tasks, suggesting that the variance within the Year 5 and Year 6 data for each of the tasks was equal. Equal variances were also observed when the School 1, 2 and 3 datasets were analysed, with the exception of the Act-Out Production task (Appendix 21).

Use of non-parametric tests

Given that all of the datasets were found to violate the assumption of normality and a substantial amount of the data violated the assumption of homogeneity of variance, non-parametric tests were utilised in the subsequent analysis conducted on the individual outcome measures (see section 3.7.2).

---

[55] Histograms Pre-test scores: Experimental groups, Appendix 16; Age groups, Appendix 17; Schools, Appendix 18. Post-test scores: TE-FMC and TE-F, Figure 5.1a, 5.1b, 5.1c (Chapter 5); Control, Appendix 19. Delayed post-test: all groups, Appendix 20.

4.2 Establishing the baseline

4.2.1 Equivalence of pre-test performance (by age and school)

Given the design of the study and allocation of pupils from different schools, classes and year groups to the TE-FMC, TE-F and Control groups (Figure 3.1), it was necessary to compare the pupils' performance at pre-test for each outcome measure by Age (i.e. Year 6 versus Year 5) and by School (1, 2, 3).

By age group (Year 5, Year 6)

Analysis conducted using a Mann Whitney U test yielded no significant differences between the Year 5 and 6 participants' performance on any of the tasks at pre-test (Appendix 22).

By school (1, 2, 3)

Analysis revealed no differences between the pre-test scores of the three schools on any of the tasks, with the exception of the Sentence Matching and Act-Out Comprehension tasks (Appendix 23). Nevertheless an examination of the descriptive statistics for these tasks indicated that although there was a small amount of divergence, all three schools were performing at chance level (Appendix 23).

Summary of analysis of baseline performance

The findings suggested that across the two age groups, as well as across the three schools, the participants were performing at a similar level and had an equivalent amount of prior knowledge of the target feature. In those instances where some divergence did occur, the participants' performance remained at chance level. Consequently the data from the two age groups within the three schools was grouped together for subsequent analysis giving three groups; TE-FMC, TE-F, and Control.

Analysis was also conducted in order to determine whether there was any effect of Age or School at post-test. No differences were found between the two age groups or between the three schools at post-test; however the results are beyond the scope of this thesis.

4.2.2 Performance on the vocabulary test

Prior to completing the outcome measures at both pre- and post-test, the learners completed a short multiple choice test of receptive vocabulary knowledge (section 3.6.6). Table 4.1 details the descriptive statistics[56] for the three groups' performance on the vocabulary test at pre- and post-test (see also Figure 4.1).

[56] Non-parametric analysis is based on the median score; therefore for all tasks each group's median score (*Mdn*) is reported. The mean (*M*) and standard deviation (*SD*) are also provided, in order to provide a more exact measure of central tendency and detail the (slight) variation between groups.

Table 4.1: Descriptive statistics for the vocabulary test

Group	N	Pre-test (k = 20)			Post-test (k = 20)		
		M	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	43	17.34	2.22	18	17.98	1.82	18
TE-F	40	16.72	2.34	17	17.48	2.35	18
Control	52	15.54	2.56	16	16.69	2.75	17
Total	135	16.46	2.50	17	17.33	2.42	18

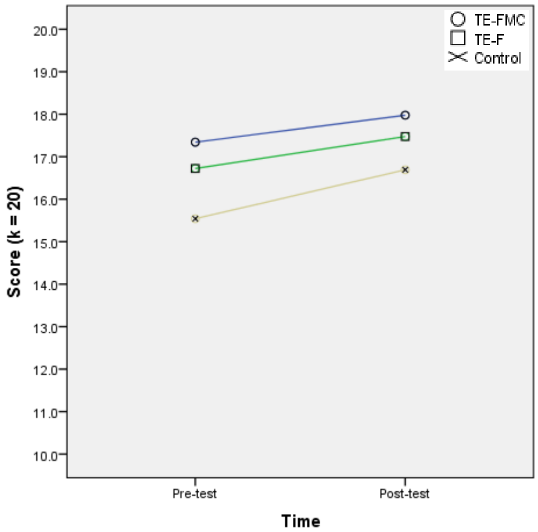


Figure 4.1: Total scores on the vocabulary test

Over time

No change in scores was found for the TE-FMC group between pre- and post-test (Wilcoxon signed rank, T = 360.000, p = .127, r = .16). In contrast a significant improvement was revealed between pre- and post-test for both the TE-F (T = 363.500, p = .022, r = .26) and Control groups (T = 806.000, p = .004, r = .29).

Between groups

A significant difference between the three groups was found at pre-test (Kruskall Wallis, $H(2)$ = 16.054, p = .001) with the TE-FMC group's scores being significantly higher than that of the Control group (p = .001, r = .40). In contrast no difference was found between the three groups at post-test ($H(2)$ = 5.622, p = .060).

Summary of vocabulary test findings

The findings suggested that at pre-test the TE-FMC group were slightly, albeit it significantly, more familiar with the vocabulary than either the TE-F or Control groups. By post-test, however, both the TE-F and Control groups' scores on the vocabulary test had improved and no difference was found between the three groups. Furthermore, the difference between the three groups' vocabulary knowledge at pre-test did not appear to result in any advantage for the TE-FMC group when completing the outcomes measures at pre-test (sections 4.3, 4.4, 4.6).

4.3 Performance on the written outcome measures

4.3.1 *Sentence Matching* task

4.3.1.1 Analysis of total scores

The descriptive statistics (Table 4.2, Figure 4.2) indicated that the TE-FMC and TE-F groups had improved at post-test and sustained this improvement at delayed post-test. In contrast there was no change in the Control group scores between pre- and post-test.

Table 4.2: Descriptive statistics for the Sentence Matching task

Group	N	Pre-test			Post-test			Delayed post-test		
		M	***SD***	***Mdn***	***M***	***SD***	***Mdn***	***M***	***SD***	***Mdn***
		Total (*k* = 24)								
TE-FMC	45	11.95	1.22	12	19.31	4.91	22	18.49	4.71	21
TE-F	38	11.97	1.17	12	19.66	4.47	21	19.08	4.92	21
Control	50	12.18	1.41	12	11.96	1.16	12	-	-	-
Total	*	12.05	1.28	12	16.56	5.42	14	18.76	4.79	21
Group	**N**	**SVO (*k* = 12)**								
TE-FMC	45	11.48	1.37	12	11.16	1.49	12	10.89	1.32	11
TE-F	38	11.76	1.32	12	11.13	1.51	12	11.16	1.60	12
Control	50	10.70	1.93	11	11.04	1.62	12	-	-	-
Total	*	11.27	1.65	12	11.11	1.53	12	11.01	1.45	12
Group	**N**	**OVS (*k* = 12)**								
TE-FMC	45	0.97	1.41	0	8.16	4.29	10	7.60	5.54	10
TE-F	38	0.61	0.79	0	8.53	4.06	10	7.92	4.68	10.5
Control	50	1.48	2.01	1	0.92	1.75	0	-	-	-
Total	*	1.06	1.57	0.3	5.54	4.99	5	7.75	4.58	10

* Pre- and post-test, *N* = 133; Delayed post-test, *N* = 83

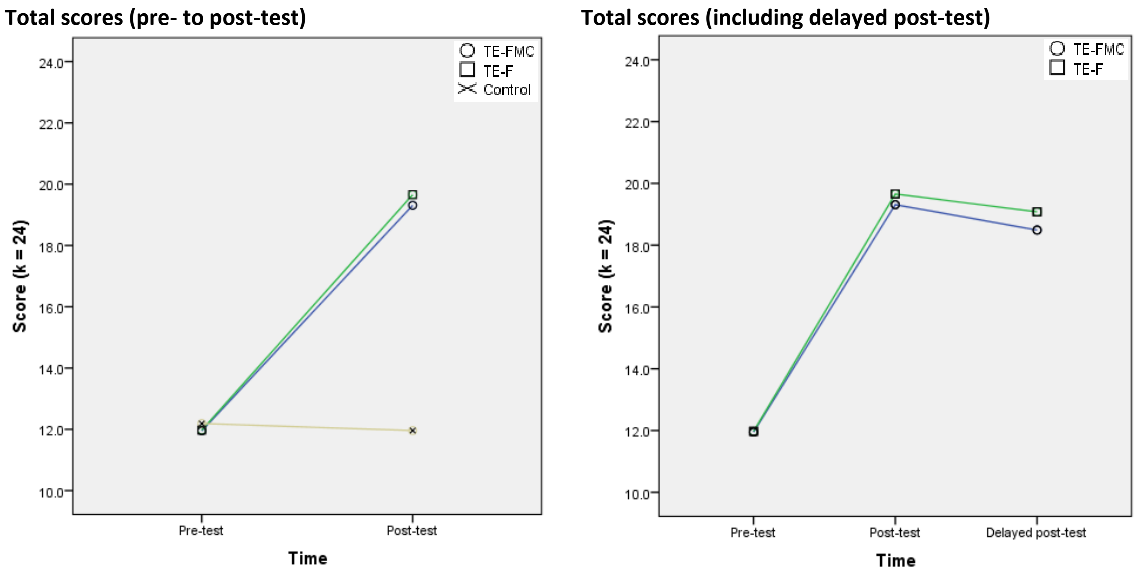


Figure 4.2: Total scores on the Sentence Matching task

Over time (pre- to post-test)

There was a significant improvement in both the TE-FMC (Wilcoxon signed-rank, $T = 894.00$, $p = .001$, $r = .58$) and TE-F groups' ($T = 621.00$, $p = .001$, $r = .58$) scores at post-test. In contrast there was no change in the Control groups' scores (Wilcoxon signed-rank, $T = 341.50$, $p = .350$, $r = -.09$).

Over time (pre-, post- to delayed post-test)

When analysing the participants' scores over all three time points (pre-, post- and delayed post-test) a significant change was found for both the TE-FMC (Friedman's ANOVA, χ^2 (2) = 49.617, $p = .001$), and TE-F groups (χ^2 (2) = 33.793, $p = .001$). Pairwise comparisons[57] revealed a significant difference between pre- and post-test for both groups (TE-FMC, $z = -6.166$, $p = .001$, $r = -.65$; TE-F, $z = -4.932$, $p = .001$, $r = -.57$). Additionally a significant difference between pre- and delayed post-test was found for both the TE-FMC group ($z = -5.534$, $p = .001$, $r = -.58$) and the TE-F group ($z = -4.531$, p = .001, $r = -.52$). In contrast no difference was found between the post-test and delayed post-test for either group (TE-FMC, $z = 0.632$, $p = .527$, $r = .07$; TE-F, $z = -4.531$, $p = .688$, $r = .05$).

Between groups (pre- and post-test)

There was no difference between the three groups at pre-test (Kruskall Wallis, $H(2) = 2.11$, $p = .348$); however there was a significant difference in the performance of the three groups at post-test ($H(2) = 60.88$, $p = .001$). Pairwise comparisons revealed that there was no difference between the TE-FMC and TE-F groups ($p = .965$, $r = .00$). The Control group's scores were significantly different from both the TE-FMC ($p = .001$, $r = .70$) and the TE-F groups ($p = .001$, $r = .69$).

Between groups (delayed post-test)

No difference was found between the TE-FMC and TE-F groups at delayed post-test (Mann Whitney, $U = 839.500$, $z = -.143$, $p = .439$, $r = -.02$), suggesting that both groups continued to perform at a similar level on this task at delayed post-test.

4.3.1.2 Analysis by test condition (SVO, OVS)

The Sentence Matching task contained test items in two different word order conditions; SVO and OVS ($k = 12$ per condition). It was expected that all participants would perform at ceiling level on the SVO sentences at pre-test; therefore any learning gains would be evident in the participants' scores for the OVS sentences (see section 3.6.3.2).

[57] Pairwise comparisons were based on an adjusted alpha level ($p = .0167$), calculated by applying a Bonferroni correction (.05 / number of comparisons made (i.e. 3)) (Field, 2009; Larson-Hall, 2010) (see section 3.7.2.2)

i) SVO test condition

Over time (pre- to post-test)

For the SVO test condition (Table 4.2, above), the initial hypothesis was correct. All three groups were performing at ceiling level on this condition at both pre- and post-test (Figure 4.3a). No significant change was found between pre- and post-test for either the TE-FMC (Wilcoxon Signed rank, T = 223.00, p = .348, r = -.08), or Control groups (T = 345.50, p = .237, r = .12). For the TE-F group, the difference between pre- and post-test approached significance (T = 62.00, p = 0.059, r = -.22), due to a decrease in scores on the SVO condition at post-test.

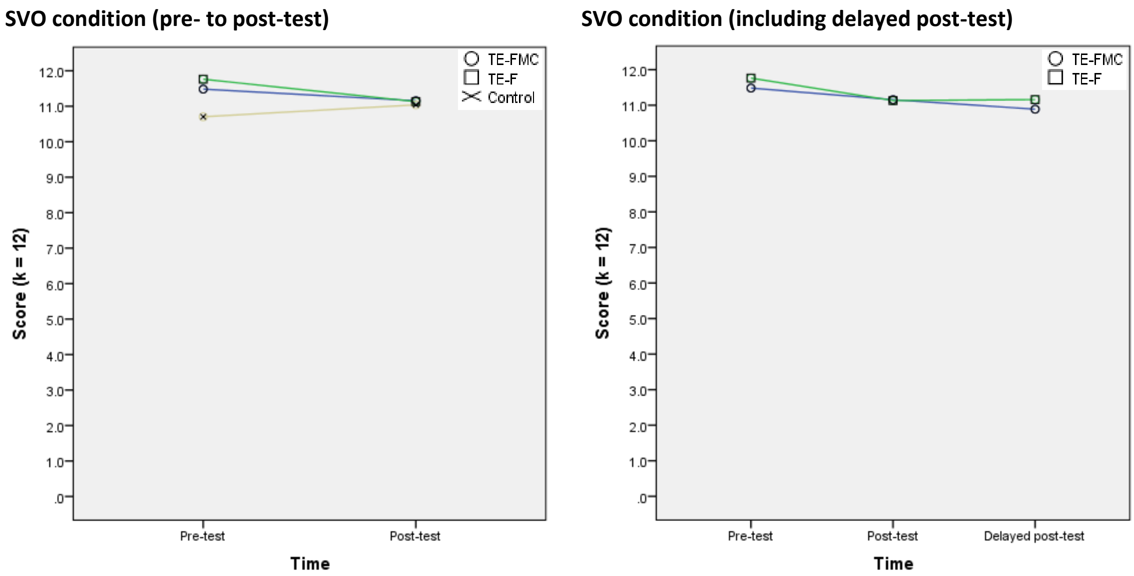


Figure 4.3a: Scores on the SVO test condition (Sentence Matching)

Over time (pre-, post- to delayed post-test)

Friedman's ANOVA revealed no change in the learners' performance on the SVO test items over the three time points for either the TE-FMC (χ^2 (2) = 4.105, p = .128) or TE-F groups (χ^2 (2) = 3.694, p = .158). The TE-FMC and TE-F groups continued to perform at ceiling level on the SVO sentences at delayed post-test (Figure 4.3a).

Between groups (pre- and post-test)

No difference was found between the three groups at post-test (Kruskall Wallis, H(2) = 0.835, p = .659). However a significant difference was found at pre-test (H(2) = 9.281, p = .010). Pairwise comparisons revealed a significant difference between the TE-F and Control groups (p = .003, r = .32), due to the TE-F group (Mdn = 12) performing slightly higher on the SVO condition than the Control group (Mdn = 11) at pre-test.

Between group (delayed post-test)

No difference was found between the TE-FMC and TE-F groups' performance at delayed post-test (Mann Whitney, U = 1010.000, z = 1.540, p = .123, r = .17).

ii) OVS test condition

Over time (pre- to post-test)

Significant changes were documented for all three groups on the OVS condition; for both the TE-FMC (T = 934.00, p = .001, r = .59) and TE-F groups (T = 660.00, p = .001, r = .59) this change reflected a significant increase in their scores on this condition. For the Control group the significant change (T = 117.50, p = .017, r = -.24) reflected a decrease in scores at post-test as observed in the descriptive statistics (Table 4.2 above, Figure 4.3b).

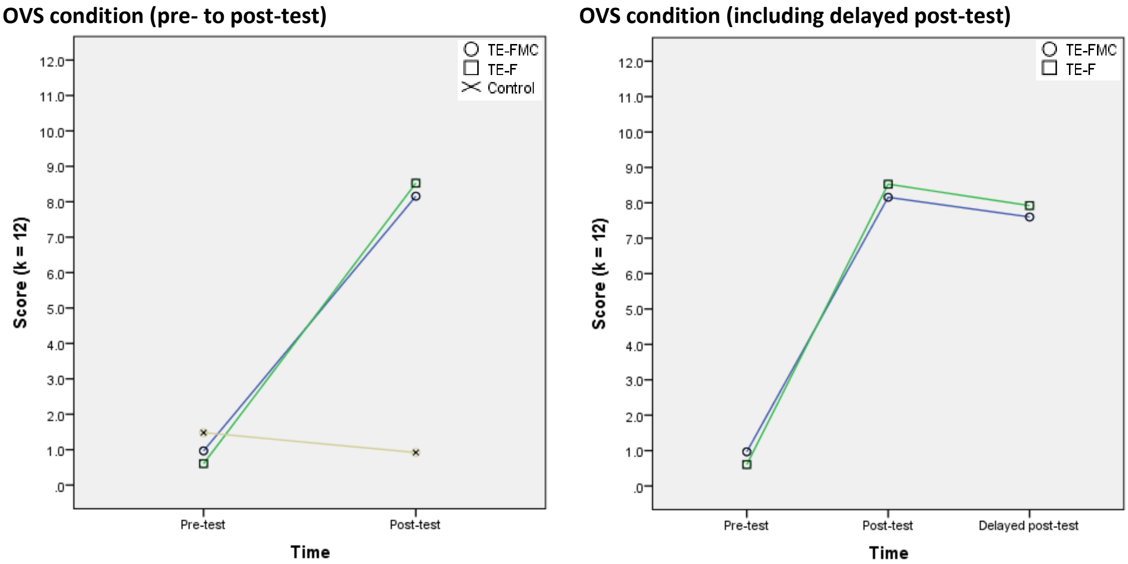


Figure 4.3b: Scores on the OVS test condition (Sentence Matching)

Over time (pre-, post- to delayed post-test)

An overall significant change over time was observed for the TE-FMC group (Friedman's ANOVA, χ^2 (2) = 45.436, p = .001), with a significant difference occurring between the learners' pre- and post-test scores (z = -6.061, p = .001, r = -.64) and pre- and delayed post-test scores (z = -4.849, p = .001, r = -.51). Between post- and delayed post-test, however, no difference was observed (z = 1.212, p = .225, r = .13). The same pattern was observed for the TE-F group (χ^2 (2) = 40.358, p = .001), with the participants' post-test (z = -5.277, p = .001, r = -.61) and delayed post-test (z = -5.047, p = .001, r = .58) scores being significantly higher than the pre-test scores, but no difference between the post- and delayed post-test (z = 0.229, p = .819, r = .03).

Between groups (pre- and post-test)

A comparison of the three groups' scores at each time point revealed no difference between the groups at pre-test (Kruskall Wallis, H(2) = 4.309, p = .116); however a significant difference was found at post-test (H(2) = 67.97, p = .001). Pairwise comparisons revealed a significant difference between the TE-FMC group and the Control group (p = .001, r = .74), and between the TE-F group and the Control group (p = .001, r = .73). No difference was found between the TE-FMC and TE-F groups (p = .932, r = .09).

Between groups (delayed post-test)

As reflected in the descriptive statistics, there was no difference between the TE-FMC and TE-F groups' performance on the OVS condition at delayed post-test (Mann Whitney, U = 924.00, z = .640, p = .522, r = .070).

4.3.1.3 Examination of effect sizes

First, the magnitude of instructional effect for both the TE-FMC and TE-F groups when compared to the Control group was large[58] (Table 4.3).

Table 4.3: Magnitude of instructional effect on the Sentence Matching task

Contrast	Group(s)	*d*
Between groups	**TE-FMC vs. Control**	2.50
	TE-F vs. Control	2.98
Pre to post	**TE-FMC**	2.40
	TE-F	2.69
	Control	-0.17
Pre to delayed post	**TE-FMC**	2.32
	TE-F	2.35

Secondly, the magnitude of change between pre- and post-test was large for both the TE-FMC and TE-F groups, reflecting the learners' significant improvement on this task. In contrast the magnitude of change for the Control group was small. Thirdly, the magnitude of change between pre- and delayed post-test remained large for both the TE-FMC and TE-F groups.

4.3.2 *Gap-fill* task

4.3.2.1 Analysis of total scores

Table 4.4 contains the descriptive statistics for the Gap-fill task (see also Figure 4.4).

Over time (pre- to post-test)

A significant improvement was documented for both the TE-FMC (Wilcoxon signed rank, T = 764.50, p = .001, r = .50) and TE-F groups (T = 598.00, p = .001, r = .53). A significant improvement was also found in the Control group scores between pre- and post-test (T = 295.50, p = .032, r = .21).

[58] $0.2 < d < 0.5$ is considered small, $0.5 < d < 0.8$ medium, $d > 0.8$ large (J. Cohen, 1988; Norris & Ortega, 2000)

Table 4.4: Descriptive statistics for the Gap-fill task

Group	N	Pre-test *M*	Pre-test *SD*	Pre-test *Mdn*	Post-test *M*	Post-test *SD*	Post-test *Mdn*	Delayed post-test *M*	Delayed post-test *SD*	Delayed post-test *Mdn*
Total (*k* = 24)										
TE-FMC	45	11.55	1.36	12	17.51	5.54	20	16.93	5.00	15
TE-F	38	11.34	2.13	12	18.24	5.52	21	17.26	5.26	16.5
Control	50	11.16	2.60	12	11.62	2.36	12	-	-	-
Total	*	11.35	2.10	12	15.50	5.48	12	17.08	5.09	16
SVO+Subj (*k* = 6)										
TE-FMC	45	5.77	0.67	6	5.29	1.16	6	5.33	1.28	6
TE-F	38	5.34	1.34	6	5.18	1.49	6	5.42	1.24	6
Control	50	5.14	1.83	6	5.42	1.60	6	-	-	-
Total	*	5.41	1.40	6	5.31	1.43	6	5.37	1.26	6
OVS+Subj (*k* = 6)										
TE-FMC	45	5.60	0.72	6	4.56	1.95	5	5.13	1.39	6
TE-F	38	5.50	1.22	6	5.00	1.64	6	5.29	1.33	6
Control	50	5.01	1.76	6	5.18	1.77	6	-	-	-
Total	*	5.38	1.34	6	4.92	1.81	6	5.21	1.36	6
SVO+Obj (*k* = 6)										
TE-FMC	45	0.05	0.21	0	4.11	2.10	5	3.24	2.50	3
TE-F	38	0.24	0.88	0	4.03	2.30	5	3.18	2.52	3
Control	50	0.50	1.54	0	0.62	1.52	0	-	-	-
Total	*	0.27	1.07	0	2.77	2.58	3	3.22	2.49	3
OVS+Obj (*k* = 6)										
TE-FMC	45	0.14	0.55	0	3.56	2.41	5	3.22	2.55	4
TE-F	38	0.26	1.16	0	4.03	2.19	5	3.37	2.50	4
Control	50	0.44	1.40	0	0.40	1.25	0	-	-	-
Total	*	0.29	1.10	0	2.50	2.56	2	3.29	2.51	4

* Pre- and post-test, *N* = 133; Delayed post-test, *N* = 83

Total scores (pre- to post-test) **Total scores (including delayed post-test)**

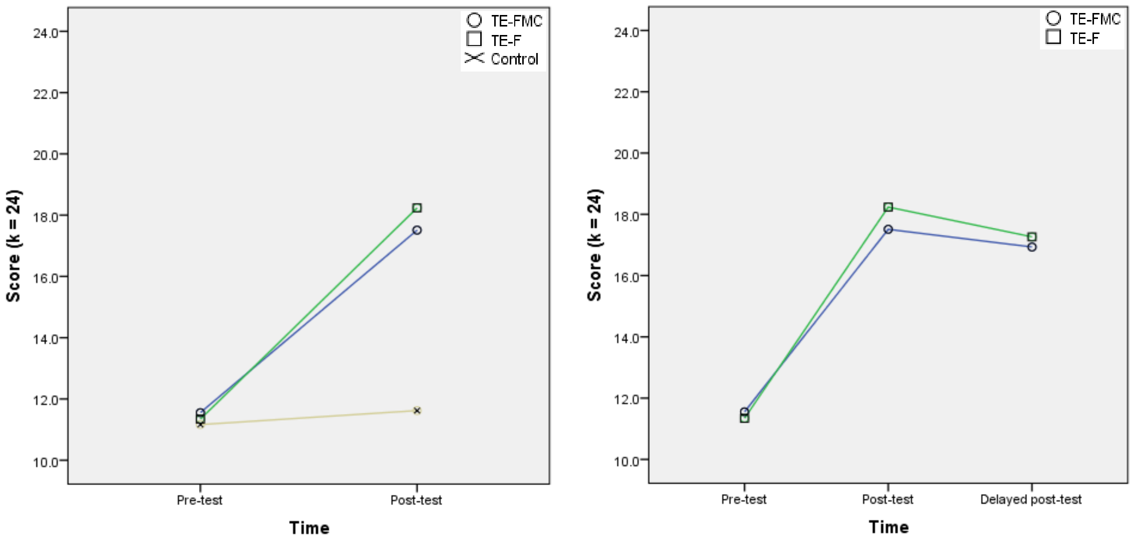


Figure 4.4: Total scores on the Gap-fill task

Over time (pre-, post- to delayed post-test)

Friedman's ANOVA yielded a significant change over time for the TE-FMC group (χ^2 (2) = 27.269, *p* = .001). Pairwise comparisons revealed a significant difference between the learners' pre- and post-test (*z* = -3.953, *p* = .001, *r* = -.42), as well as delayed post-test scores (*z* = -4.427, *p* = .001, *r* = -.47), but no change between post- and delayed post-test (*z* = -0.474, *p* = .635, *r* = .05). With regard to the TE-F group performance, a significant change over time was also found (χ^2 (2) = 32.294, *p* = .001) with a significant difference between pre- and post-test (*z* = -4.474, *p* = .001, *r* = -.51) and pre- and delayed post-test (*z* = -4.818, *p* = .001, *r* = -.55), but no difference between post- and delayed post-test (*z* = -0.344, *p* = .731, *r* = -.04).

Between groups (pre- and post-test)

The results of the Kruskall Wallis test revealed that there was no difference between the three groups at pre-test (H(2) = 0.643, *p* = .725). In contrast a significant difference between groups was observed at post-test (H(2) = 36.043, *p* = .001). Pairwise comparisons revealed no difference between the TE-FMC and TE-F groups' scores at post-test (*p* = .364 *r* = -.10). However the Control group scores were found to be significantly different to both the TE-FMC (*p* = .001, *r* = .48) and the TE-F groups' scores (*p* = .001, *r* = .58).

Between groups (delayed post-test)

At delayed post-test, a Mann Whitney U test revealed no difference between the two groups (*U* = 866.000, *z* = .102, *p* = .919, *r* = .01).

4.3.2.2 Analysis by test condition (SVO+Subj, OVS+Subj, SVO+Obj, OVS+Obj)

The Gap-fill task contained test sentences in four conditions; SVO+Subj (SVO word order; subject missing), OVS+Subj, SVO+Obj, and OVS+Obj (*k* = 6 for each condition; section 3.6.3.3). For the missing subject conditions (SVO+Subj, OVS+Subj) it was anticipated that all pupils would perform at ceiling level at pre-test. During the pre-teaching phase all of the pupils, across all three groups, had been introduced to the masculine nominative article ***der*** (subject); therefore it was expected that at pre-test the pupils would use this article (***der***) indiscriminately across all of the test conditions. The critical SVO+Obj and OVS+Obj conditions would demonstrate whether the pupils were correctly able to produce the target feature, the accusative article (***den***), in SVO and OVS sentences. The descriptive statistics for the four test conditions can be seen in Table 4.4 (above).

i) SVO+Subj condition

Over time (pre- to post-test)

No change over time was found for either the TE-F (Wilcoxon signed rank, T = 88.00, p = .517, r = -.07) or Control groups (T = 97.50, p = .123, r = .15) on the SVO+Subj condition. For the TE-FMC group, however, a slight, yet significant, decrease in scores was observed between pre- and post-test (T = 47.00, p = .015, r = -.26; Figure 4.5a).

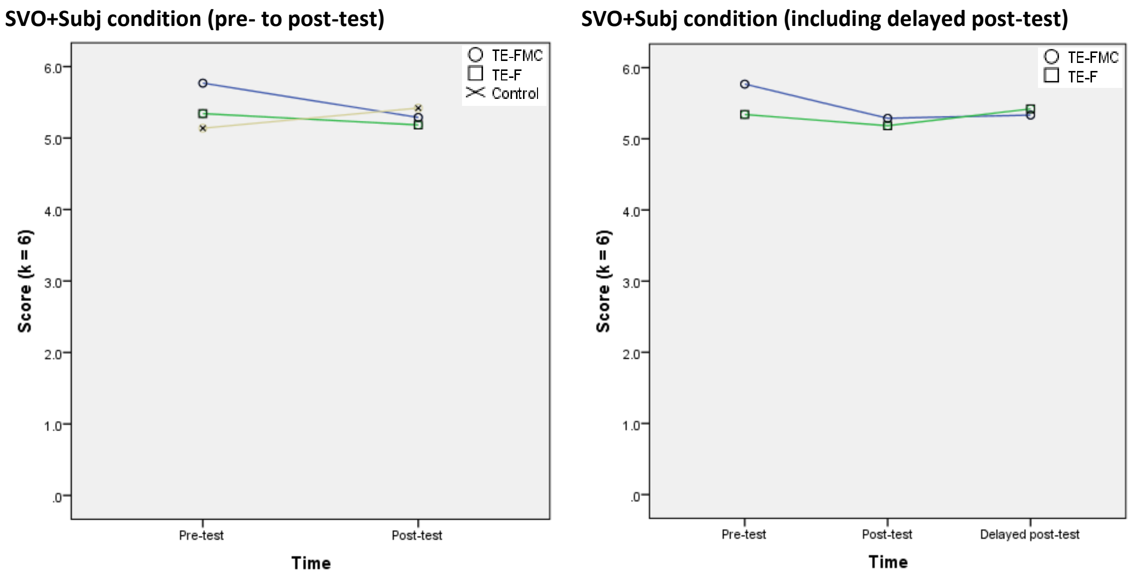


Figure 4.5a: Scores on the SVO+Subj test condition (Gap-fill)

Over time (pre-, post- to delayed post-test)

With regard to the TE-FMC and TE-F groups' performance over the three time points (Figure 4.5a), Friedman's ANOVA revealed no change over time on the SVO+Subj condition for either group (TE-FMC, χ^2 (2) = 3.205, p = .201; TE-F, χ^2 (2) = 0.775, p = .679).

Between groups (pre to post-test)

Despite the slight decrease in the TE-FMC group's scores at post-test, no difference was found between the three groups at either pre-test (Kruskall Wallis, H(2) = 2.75, p = .253) or post-test (H(2) = 3.96, p = .138).

Between groups (delayed post-test)

A Mann Whitney U test revealed no difference between the TE-FMC and TE-F groups on the SVO+Subj condition at delayed post-test (U = 859.000, z = 0.048, p = .962, r = .01).

ii) OVS+Subj condition

Over time (pre- to post-test)

For the OVS+Subj condition (Table 4.4 above, Figure 4.5b), no difference was found between pre- and post-test for either the Control group (Wilcoxon signed rank, T = 137.00, p = .449, r = .08) or the TE-F group (T = 46.00, p = .079, r = -.20), although the result of the Wilcoxon signed-rank test for the TE-F group was approaching significance. A significant decrease was observed in the TE-FMC group's scores on the OVS+Subj test items at post-test (T = 68.50, p = .004, r = .31).

OVS+Subj condition (pre- to post-test) **OVS+Subj condition (delayed post-test)**

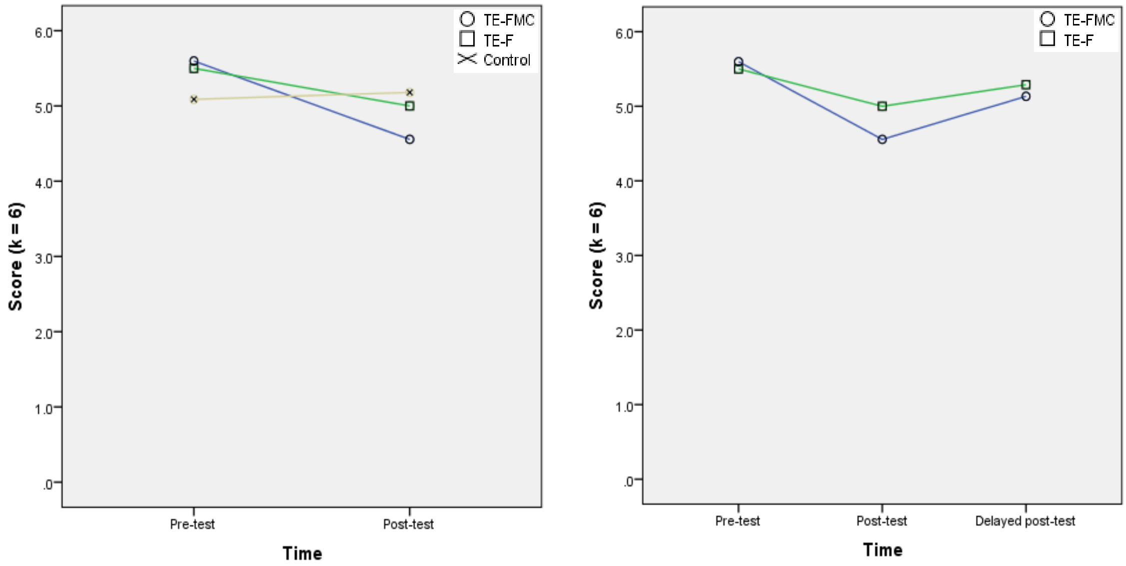


Figure 4.5b: Scores on the OVS+Subj test condition (Gap-fill)

Over time (pre-, post- to delayed post-test)

Friedman's ANOVA yielded no change over the three time points for the TE-F group (χ^2 (2) = 3.840, p = .147). The analysis of the TE-FMC group's scores, however, revealed a significant overall change in their performance on the OVS+Subj condition (χ^2 (2) = 6.743, p = 0.034). The difference between pre- and post-test approached significance[59] (z = 2.055, p = .040, r = .22). No difference was found between pre- and delayed post-test (z = 0.949, p = .343, r = .10) or post- and delayed post-test (z = -1.107, p = .268, r = .12). Along with the descriptive statistics (Figure 4.5b), this indicated that the overall significant change reflected a slight decrease in the TE-FMC learners' scores between pre- and post-test, as observed above.

Between groups (pre- and post-test)

A Kruskall Wallis test revealed no difference between the three groups at pre-test (H(2) = 1.87, p = .393). At post-test, however, a significant difference was observed (H(2) = 6.507, p = .039). Pairwise comparisons revealed that the Control group's scores at post-test were significantly higher than the TE-FMC group's scores (p = .011, r = -.26; see Table 4.4 and Figure 4.5b).

Between groups (delayed post-test)

No difference was found between the TE-FMC and TE-F groups' performance on the OVS+Subj condition at delayed post-test (Mann Whitney, U = 933.000, z = 0.824, p = .410, r = .09).

[59] Pairwise comparisons were based on an adjusted alpha level (p = .0167) (see section 3.7.2.2)

iii) SVO+Obj condition

With regards to the two test conditions which required the pupils to 'fill in' the object of the sentence (SVO+Obj, OVS+Obj), a greater amount of divergence was observed between the three groups (Table 4.4).

Over time (pre- to post-test)

Analysis of the TE-FMC (Wilcoxon signed rank, T = 861.00, p = .001, r = .59) and TE-F groups' scores (T = 526.00, p = .001, r = .57) revealed a significant change between pre- and post-test with both groups scoring significantly higher at post-test (Figure 4.5c). In contrast, no change was observed in the Control group performance between pre- and post-test (T = 37.50, p = .684, r = .04).

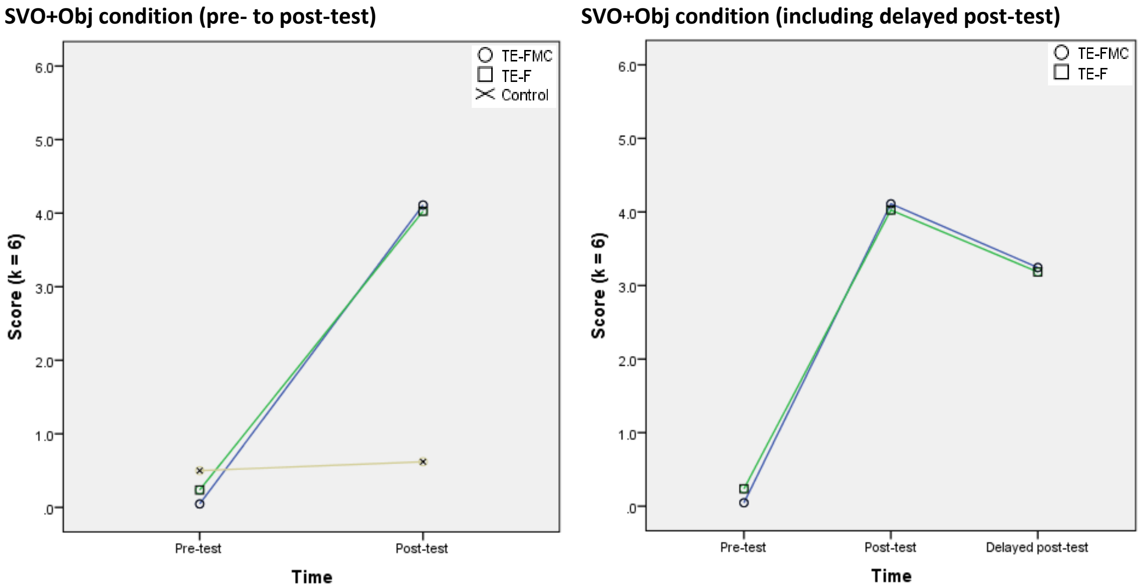


Figure 4.5c: Scores on the SVO+Obj test condition (Gap-fill)

Over time (pre-, post- to delayed post-test)

Friedman's ANOVA revealed a significant change in the participants' performance across the three time points for both the TE-FMC (χ^2 (2) = 53.737, p = .001) and TE-F groups (χ^2 (2) = 38.032, p = .001). Both groups' scores were found to be significantly higher at post-test (TE-FMC, z = -6.430, p = .001, r = -.68; TE-F, z = -5.105, p = .001, r = -.59) and at delayed post-test (TE-FMC, z = -4.954, p = .001, r = -.52; TE-F, z = -4.531, p = .001, r = -.52) than at pre-test, whereas no difference was observed between post-test and delayed post-test (TE-FMC, z = 1.476, p = .140, r = .16; TE-F, z = 0.574, p = .566, r = .07).

Between groups (pre- and post-test)

A Kruskall Wallis test revealed no difference between the three groups at pre-test (H(2) = 1.04, p = .595). However the three groups were found to be performing differently at post-test (H(2)

= 58.31, *p* = .001), with the Control group scoring significantly lower than either the TE-F (*p* = .001, *r* = .67) or TE-FMC group (*p* = .001, *r* = .69).

Between groups (delayed post-test)

No difference was found between the TE-FMC and TE-F groups at delayed post-test on the SVO+Obj condition (Mann Whitney, *U* = 847.000, *z* = -0.075, *p* = .940, *r* = -.01).

iv) OVS+Obj condition

Over time (pre- to post-test)

A significant change over time was observed for the TE-FMC (Wilcoxon signed rank, *T* = 663.50, *p* = .001, *r* = .55) and TE-F groups (*T* = 558.00, *p* = .001, *r* = .57) between pre- and post-test on the OVS+Obj condition (Figure 4.5d). No difference was observed in the Control group's scores (*T* = 29.50, *p* = .838, *r* = .02).

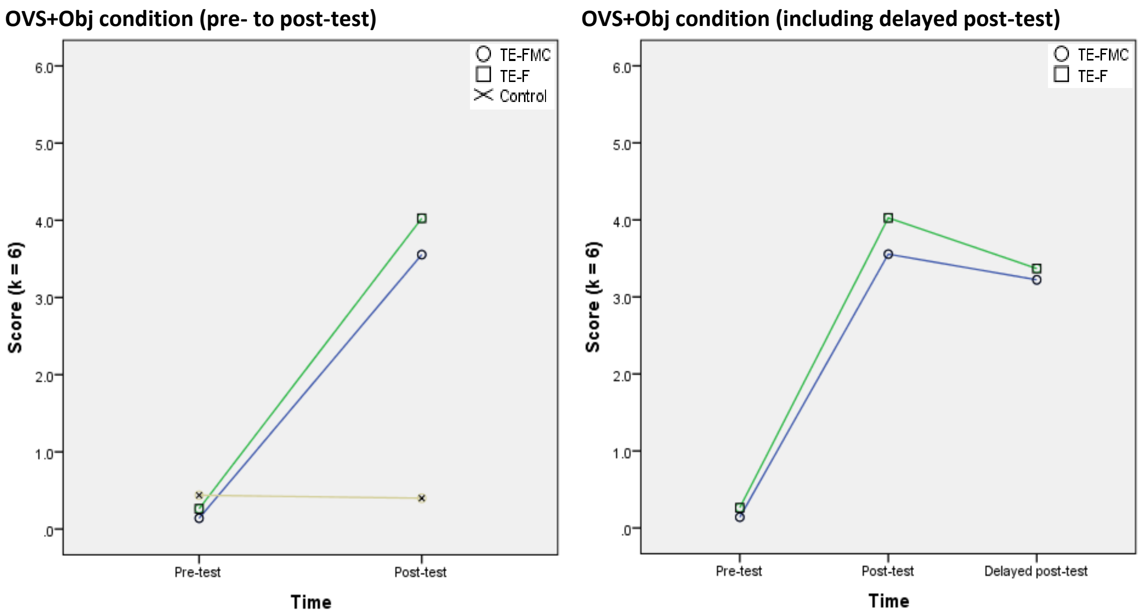


Figure 4.5d: Scores on the OVS+Obj test condition (Gap-fill)

Over time (pre-, post- to delayed post-test)

The TE-FMC group's scores were found to significantly change over the three time points (Friedman's ANOVA, χ^2 (2) =47.091, *p* = .001), due to a significant increase in the learners' scores between pre- and post-test (*z* = -5.376, *p* = .001, *r* = -.57) and pre- and delayed post-test (*z* = -4.743, *p* = .001, *r* = -.50). There was no change in the TE-FMC group's performance between post- and delayed post-test (*z* = 0.632, *p* = .527, *r* = .07). Likewise the TE-F group's performance significantly improved over time (χ^2 (2) = 42.017, *p* = .001). Pairwise comparisons found that the scores at post-test (*z* = -5.277, *p* = .001, *r* = -.61) and at delayed post-test (*z* = -4.531, *p* = .001, *r* = -.52) were significantly higher than at pre-test, but that there was no difference between post- and delayed post-test (*z* = 0.746, *p* = .456, *r* = -.09).

Between groups (pre- and post-test)

All three groups were performing at an equivalent level at pre-test (Kruskall Wallis, $H(2) = 1.135$, $p = .567$). However at post-test there was a significant difference between groups ($H(2) = 57.03$, $p = .001$), with the TE-FMC ($p = .001$, $r = .63$) and TE-F groups ($p = .001$, $r = .71$) significantly outperforming the Control group.

Between groups (delayed post-test)

A Mann Whitney U test revealed no difference between the performance of the TE-FMC and TE-F groups at delayed post-test ($U = 894.500$, $z = .372$, $p = .710$, $r = .04$).

4.3.2.3 Examination of effect sizes

Table 4.5 details the effect sizes calculated based on the three groups' performances on the Gap-fill task. The between group effect size effect sizes revealed a larger effect for the TE-FMC and TE-F groups over the Control group at post-test. In addition, the magnitude of change was large for both the TE-FMC and TE-F groups, indicating that both groups had made (post-test), and sustained (delayed post-test), gains of over one and a half standard deviations from pre-test. In contrast the change in the Control group's scores was small.

Table 4.5: Magnitude of instructional effect on the Gap-fill task

Contrast	Group(s)	*d*
Between groups	**TE-FMC vs. Control**	1.52
	TE-F vs. Control	1.78
Pre to post	**TE-FMC**	1.73
	TE-F	1.79
	Control	0.19
Pre to delayed post	**TE-FMC**	1.69
	TE-F	1.60

4.3.3 Summary of the findings for the written outcome measures

The analyses of the written task data revealed that both the TE-FMC and TE-F groups had made a significant, equivalent level of improvement on the untimed, written tasks following their respective interventions, in terms of their accuracy in both comprehending (Sentence Matching) and producing (Gap-fill) the target feature (***den***). Further both groups sustained this improvement across the delayed post-test. In contrast no improvement was observed for the Control group on either task.

For the Sentence Matching task the analysis by test condition revealed that the TE-FMC and TE-F learners' overall improvement was due to an improvement in their comprehension of the OVS test items.

For the Gap-fill task the analysis of the SVO+Obj and OVS+Obj conditions indicated that the overall improvement in the TE-FMC and TE-F groups' scores at post-test could be attributed to an increase in the number of correct responses to those sentences requiring production of ***den***. In addition, although all three groups were performing at ceiling level on the SVO+Subj and OVS+Subj conditions, a slight decrease was observed in the TE-FMC groups' performance on these conditions at post-test. This finding indicated that to a certain extent the TE-FMC group were overgeneralising their use of the 'new' article ***den*** to the subject conditions (see section 6.2.2.3 for discussion).

4.4 Performance on the oral outcome measures

4.4.1 *Act-Out Comprehension* task

4.4.1.1 Analysis of total scores

Table 4.6 details the descriptive statistics for the Act-Out Comprehension task (also Figure 4.6):

Table 4.6: Descriptive statistics for the Act-Out Comprehension task

Group	N	Total (*k* = 18)								
		Pre-test			Post-test			Delayed post-test		
		M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
TE-FMC	45	9.29	0.69	9	12.67	3.78	12	12.91	3.62	13
TE-F	41	9.19	0.98	9	12.42	3.20	12	12.34	3.26	12
Control	46	8.65	1.04	9	9.17	0.95	9	-	-	-
Total	*	9.04	0.95	9	11.37	3.30	10	12.64	3.45	12
Group	**N**	**SVO (*k* = 9)**								
TE-FMC	45	8.80	0.40	9	8.73	0.58	9	8.67	0.64	9
TE-F	41	8.54	0.81	9	8.63	0.62	9	8.63	0.66	9
Control	46	8.35	1.06	9	8.78	0.42	9	-	-	-
Total	*	8.56	0.82	9	8.72	0.54	9	8.65	0.65	9
Group	**N**	**OVS (*k* = 9)**								
TE-FMC	45	0.49	0.69	0	3.93	3.69	3	4.24	3.59	4
TE-F	41	0.65	0.94	0	3.78	3.38	3	3.71	3.38	4
Control	46	0.26	0.61	0	0.39	1.00	0	-	-	-
Total	*	0.46	0.76	0	2.65	3.34	1	3.99	3.48	4

* Pre- and post-test, *N* = 132; Delayed post-test, *N* = 86

Over time (pre- to post-test)

A significant improvement was found for both the TE-FMC (Wilcoxon signed rank, T = 631.00, p = .001, r = .45) and TE-F groups (T = 538.00, p = .001, r = .51) between pre- and post-test. Unexpectedly, analysis also revealed a significant change in the Control group scores between pre- and post-test (T = 286.00, p = .017, r = .25). A closer examination of the descriptive statistics for the Control group suggested that despite this small, yet significant, improvement, the Control group were still performing at chance level at post-test (*Mdn* = 9), and therefore this significant change did not reflect at improvement in the Control groups' comprehension of the target feature (see section 4.4.1.2).

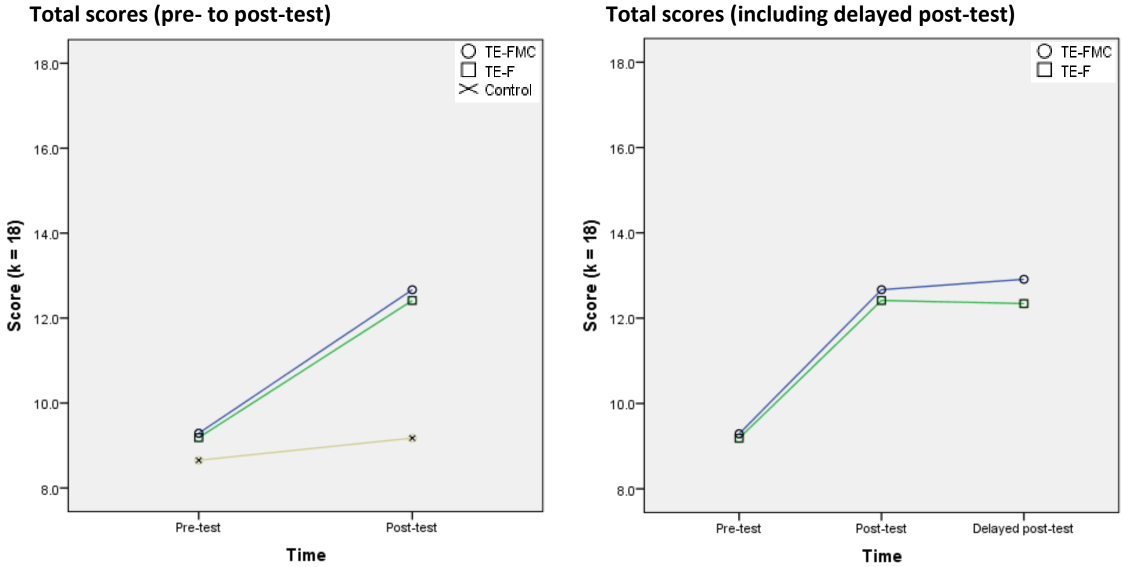


Figure 4.6: Total scores on the Act-Out Comprehension task

Over time (pre-, post- to delayed post-test)

The TE-FMC group's scores changed significantly over the three time points (Friedman's ANOVA, χ^2 (2) = 17.792, *p* = .001). Both the post- (*z* = -3.004, *p* = .001, *r* = -.32) and delayed post-test scores (*z* = -3.479, *p* = .003, *r* = -.37) were significantly higher than the pre-test scores; however there was no difference between post- and delayed post-test (*z* = -0.474, *p* = .635, *r* = -.05). Similarly there was an overall significant change in the TE-F group's scores (χ^2 (2) = 29.429, *p* = .001), with a significant change between pre- and post-test (*z* = -4.362, *p* = .001, *r* = -.48) and pre- and delayed post-test (*z* = -4.086, *p* = .001, *r* = -.45), yet no difference between post- and delayed post-test (*z* = .276, *p* = .782, *r* = .03).

Between groups (pre- and post-test)

There was a significant difference between the three groups at both pre-test (Kruskall Wallis, *H*(2) = 11.97, *p* = .003) and post-test (*H*(2) = 36.91, *p* = .001). Pairwise comparisons revealed that at pre-test the Control group scores were significantly lower than the TE-FMC group scores (*p* = .001, *r* = .35) and the difference between the Control and TE-F groups' scores at pre-test was approaching significance[60] (*p* = .024, *r* = .24). At post-test, pairwise comparisons revealed an even starker difference between the Control group and both the TE-FMC (*p* = .001, *r* = .55) and the TE-F groups (*p* = .001, *r* = .56).

[60] Pairwise comparisons were based on an adjusted alpha level (*p* = .0167) (see section 3.7.2.2)

Between groups (delayed post-test)

A Mann Whitney U test revealed no difference between the TE-FMC and TE-F groups at delayed post-test (U = 818.000, z = -.917, p = .359, r = -.10).

4.4.1.2 Analysis by test condition (SVO, OVS)

The Act-Out Comprehension task contained test sentences in two conditions; SVO (k = 9) and OVS (k = 9) word order. As with the Sentence Matching task (section 4.3.1), an improvement in the participants comprehension of the OVS sentences at post-test would demonstrate that the learners were correctly interpreting the target feature, whereas it was expected that the all of the learners would perform at ceiling level on the SVO test items from pre-test. Table 4.6 (above) details the descriptive statistics for the two test conditions.

i) SVO condition

Over time (pre- to post-test)

As hypothesised, the learners were performing at ceiling level on the SVO test items from pre-test (Table 4.6 above, Figure 4.7a). No change between pre- and post-test was found in the TE-FMC (Wilcoxon signed rank, T = 30.00, p = .439, r = -.08) or TE-F groups' scores (T = 130.00, p = .598, r = .06) on this test condition. In contrast, a significant change was found for the Control group performance (T = 178.00, p = .004, r = .30); the Control group's scores on the SVO sentences increased between pre- and post-test.

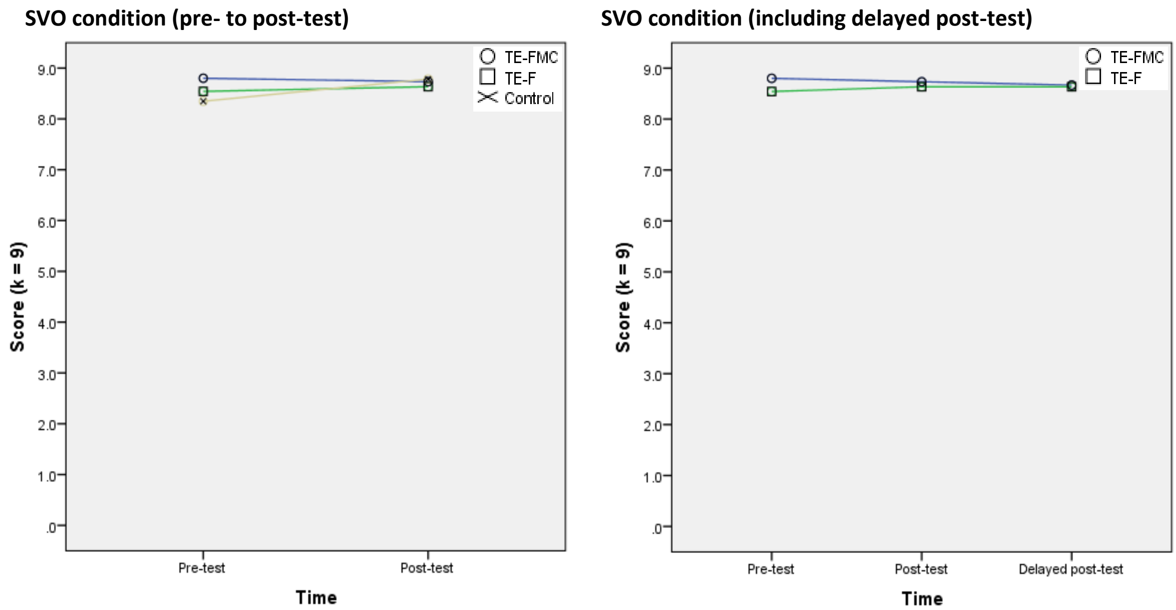


Figure 4.7a: Scores on the SVO test condition (Act-Out Comprehension

Over time (pre-, post- to delayed post-test)

Analysis via Friedman's ANOVA found no change over the three time points for either the TE-FMC (χ^2 (2) = 1.069, p = .586) or TE-F group (χ^2 (2) = 0.644, p = .725) on the SVO test items.

Between groups (pre- and post-test)

A Kruskall Wallis test revealed no difference between the three groups at post-test ($H(2)$ = 1.24, p = .539). At pre-test, however, the difference was approaching significance ($H(2)$ = 5.14, p = .072). This finding, taken together with the significant improvement in the Control group's scores between pre- and post-test, indicated that at pre-test the Control group's performance was marginally lower than that of the TE-FMC and TE-F groups; however by post-test the three groups were performing at an equivalent level.

Between groups (delayed post-test)

No difference was found between the TE-FMC and TE-F groups on this condition at delayed post-test (Mann Whitney, U = 889.00, z = -0.374, p = .708, r = -.04).

ii) OVS condition

Over time (pre- to post-test)

For the OVS test items (Table 4.6 above, Figure 4.7b), the TE-FMC (Wilcoxon signed rank, T = 603.00, p = .001, r = .45) and TE-F groups (T = 518.00, p = .001, r = .47) were found to significantly improve at post-test, whereas no change was found in the Control group scores (T = 61.00, p = .244, r = .12).

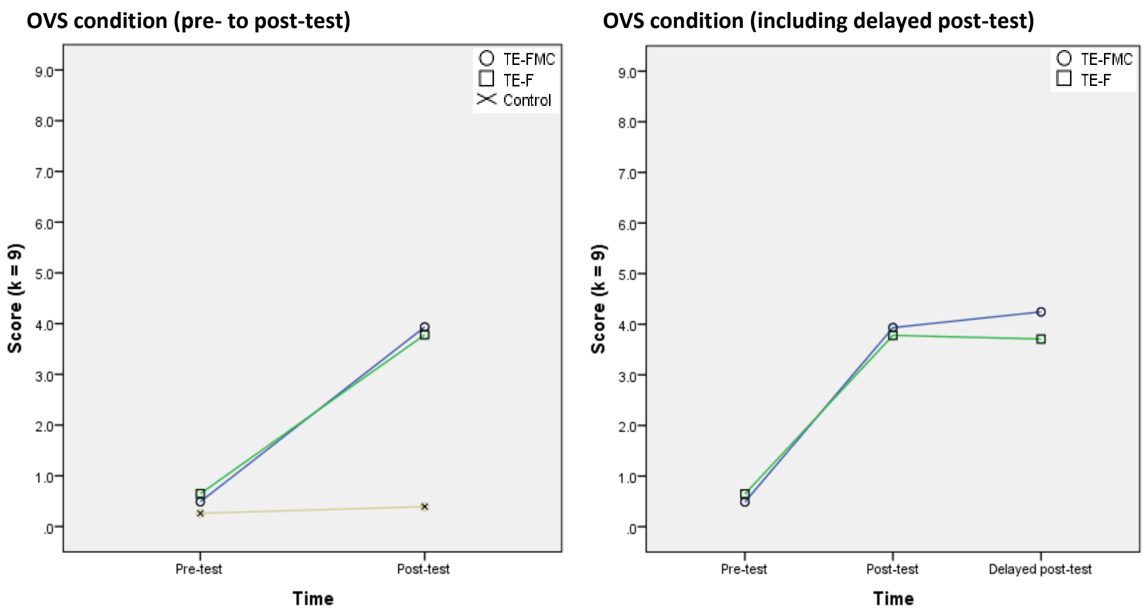


Figure 4.7b: Scores on the OVS test condition (Act-Out Comprehension)

Over time (pre-, post- to delayed post-test)

Friedmans' ANOVA yielded a significant change in both the TE-FMC (χ^2 (2) = 14.504, p = .001) and TE-F groups' (χ^2 (2) = 15.250, p = .001) scores over the three time points (Figure 4.7b). For both groups, pairwise comparisons revealed a significant difference between pre- and post-test (TE-FMC, z = -2.530, p = .011, r = -.27; TE-F, z = -3.092, p = .002, r = -.34) as well as pre- and delayed post-test (TE-FMC, z = -3.162, p = .002, r = -.33; TE-F, z = -2.871, p = .004, r = -.32) and no difference between post- and delayed post-test (TE-FMC, z = -0.632, p = .527, r = -.07; TE-F, z = .221, p = .825, r = .02).

Between groups (pre- and post-test)

At pre-test the difference between the three groups was found to be approaching significance, (Kruskall Wallis, $H(2)$ = 5.829, p = .054), reflecting the Control group's slightly lower scores (Figure 4.7b). At post-test a significant difference was found between the three groups ($H(2)$ = 35.15, p = .001), with both the TE-FMC (p = .001, r = .55) and TE-F (p = .01, r = .53) groups outperforming the Control group.

Over time (pre-, post- to delayed post-test)

At delayed post-test there was no difference between the TE-FMC and TE-F groups' scores on the OVS condition (Mann Whitney, U = 807.500, z = -1.011, p = .312, r = -.11).

4.4.1.3 Examination of effect sizes

A large instructional effect was observed for both the TE-FMC and TE-F groups over the Control group at post-test (Table 4.7). Further the magnitude of change between pre- and post-test and pre- and delayed post-test was large for both the TE-FMC and TE-F groups; whereas a medium effect was observed for the Control group. Notably the analysis of the Control group's performance revealed that this effect was due to an improvement in the learners' comprehension of SVO sentences, rather than the critical OVS test items.

Table 4.7: Magnitude of instructional effect on the Act-Out Comprehension task

Contrast	Group(s)	*d*
Between groups	**TE-FMC vs. Control**	1.49
	TE-F vs. Control	1.62
Pre to post	**TE-FMC**	1.51
	TE-F	1.55
	Control	0.52
Pre to delayed post	**TE-FMC**	1.68
	TE-F	1.49

4.4.2 *Act-Out Production* task

4.4.2.1 Analysis of total scores

The descriptive statistics for the Act-Out Production task are presented in Table 4.8 (see also Figure 4.8).

Table 4.8: Descriptive statistics for the Act-Out Production task

Group	N	Pre-test *M*	Pre-test *SD*	Pre-test *Mdn*	Post-test *M*	Post-test *SD*	Post-test *Mdn*	Delayed post-test *M*	Delayed post-test *SD*	Delayed post-test *Mdn*
Total (*k* = 24)										
TE-FMC	45	9.76	3.54	12	17.24	5.67	17	18.16	5.63	20
TE-F	41	10.44	3.18	12	16.24	6.18	16	16.93	6.49	17
Control	46	9.87	3.40	12	10.02	3.24	12	-	-	-
Total	*	10.01	3.37	12	14.42	6.05	12	17.57	6.05	19.5
Subject (*k* = 12)										
TE-FMC	45	9.76	3.54	12	11.04	2.46	12	11.47	1.32	12
TE-F	41	10.29	3.27	12	10.93	2.69	12	10.66	3.28	12
Control	46	9.87	3.40	12	10.02	3.24	12	-	-	-
Total	*	9.96	3.39	12	10.65	2.84	12	11.08	2.47	12
Object (*k* = 12)										
TE-FMC	45	0.00	0.00	0	6.20	5.81	8	6.69	5.31	9
TE-F	41	0.15	0.69	0	5.32	5.19	4	6.27	5.26	9
Control	46	0.00	0.00	0	0.00	0.00	0	-	-	-
Total	*	0.05	3.89	0	3.77	5.08	0	6.49	5.26	9

* Pre- and post-test, *N* = 132; Delayed post-test, *N* = 86

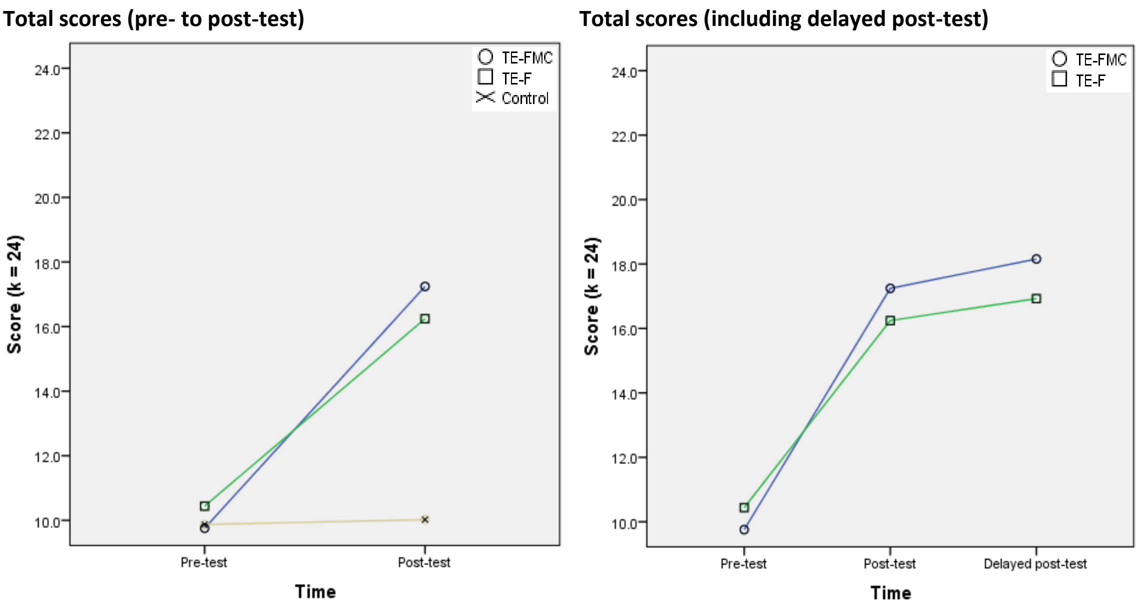


Figure 4.8: Total scores on the Act-Out Production task

Over time (pre- to post-test)

Analysis revealed a significant change in both the TE-FMC (Wilcoxon signed rank, *T* = 697.00, *p* = .001, *r* = .55), and TE-F groups' (*T* = 542.50, *p* = .001, *r* = .52) scores between pre- and post-test. No change over time was observed for the Control group (*T* = 205.00, *p* = .451 *r* = .08).

Over time (pre-, post- to delayed post-test)

Friedman's ANOVA revealed a significant change between pre-, post- and delayed post-test in both the TE-FMC (χ^2 (2) = 53.035, *p* = .001) and TE-F groups' (χ^2 (2) = 33.939, *p* = .001) scores over the three time points. Both groups scored significantly higher at post-test than at pre-test

(TE-FMC, z = -5.112, p = .001, r = -.54; TE-F, z = -4.141, p = .001, r = -.46), as well as higher at delayed post-test than at pre-test (TE-FMC, z = -5.956, p = .001, r = -.63; TE-F, z = -4.804, p = .001, r = -.53). Between post- and delayed post-test no change was found for either group (TE-FMC, z = -0.843, p = .399; r = -.09, TE-F, z = -0.663, p = .508, r = -.07).

Between groups (pre- and post-test)

No difference was found between the three groups' scores at pre-test (Kruskall Wallis, $H(2)$ = 1.37, p = .504). At post-test, however, the Control group scores were found to be significantly different ($H(2)$ = 41.20, p = .001), to both the TE-FMC (p = .001, r = .60) and the TE-F groups' (p = .001, r = .57) scores. There was no difference between the TE-FMC and TE-F groups at post-test and both groups outperformed the Control group at this time point (Figure 4.8).

Between groups (delayed post-test)

An examination of the descriptive statistics (Table 4.8, Figure 4.8) suggested that the TE-FMC group may have been outperforming the TE-F group to a certain extent at delayed post-test (TE-FMC, Mdn = 20; TE-F, Mdn = 17). Nevertheless analysis found no difference between the two groups (Mann Whitney, U = 805.000, z = -1.033, p = .301, r = -.11).

4.4.2.2 Analysis by test condition (Subject, Object)

The Act-Out Production task (section 3.6.4.2) required the learners to produce full sentences (n = 12) to describe visual stimuli (i.e. the monkey chases the tiger). Consequently the learners were required to correctly produce both the article ***der*** (subject) and ***den*** (object). Notably, all three groups were performing below ceiling level on the subject test items at pre-test (Table 4.8 above, Figure 4.9a).

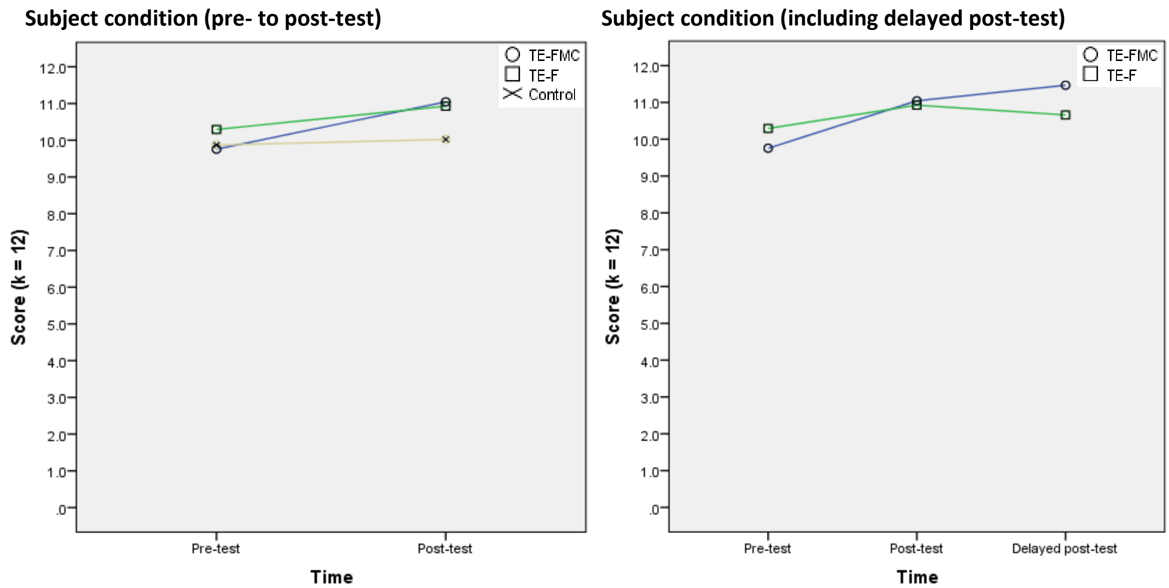


Figure 4.9a: Scores on the Subject test condition (Act-Out Production)

i) Subject condition

Over time (pre- to post-test)

There was no significant change between pre- and post-test for either the TE-F (Wilcoxon signed rank, T = 118.50, p = .342, r = .10) or Control groups (T = 205.00, p = .451, r = .08) on the subject condition. In contrast, there was a significant improvement in the TE-FMC group's scores between pre- and post-test (T = 248.50, p = .001, r = .36) (Table 4.8 above, Figure 4.9a).

Over time (pre-, post- to delayed post-test)

Friedman's ANOVA revealed a significant change over time for the TE-FMC group (χ^2 (2) = 11.341, p = .003); however pairwise comparisons revealed no difference between any of the individual time points, although a comparison of the learners' pre- and delayed post-test scores was approaching significance[61], (z = -2.214, p =.027, r = .23). As can be seen in Figure 4.9a, the TE-FMC group's performance on the subject condition was lower at pre-test than at post- and delayed post-test. The descriptive statistics (Table 4.8) indicated that there was greater accuracy, coupled with a lower level of variation, in the TE-FMC group's performance at delayed post-test when compared to pre-test. With regard to the TE-F group's performance on the subject condition, no change over time was found (χ^2 (2) = 1.740, p = .419) (see also Figure 4.9a).

Between groups (pre- and post-test)

No difference was found between the three groups at pre- (Kruskall Wallis, H(2) = 0.989, p = .610) or post-test (H(2) = 4.656, p = .097), although at post-test the difference was approaching significance. This approaching significant difference could be accounted for by the significant improvement in the TE-FMC group's scores on the subject condition (Table 4.8, Figure 4.9a).

Between groups (delayed post-test)

There was no difference between the TE-FMC and TE-F groups' performance at delayed post-test (U = 934.000, z = 0.132, p = .895, r = .01).

ii) Object condition

Over time (pre- to post-test)

An examination of the change in pupil scores found that both the TE-FMC (Wilcoxon signed rank, T = 496.00, p = .001, r = .52) and TE-F groups (T = 372.00, p = .001, r = .49) made substantial improvement between pre- and post-test, whereas no improvement was found in the Control group performance (T = 0.00, p = 1.000) (Table 4.8 above, Figure 4.9b).

[61] For pairwise comparisons, adjusted alpha level was p = .0167 (section 3.7.2.2)

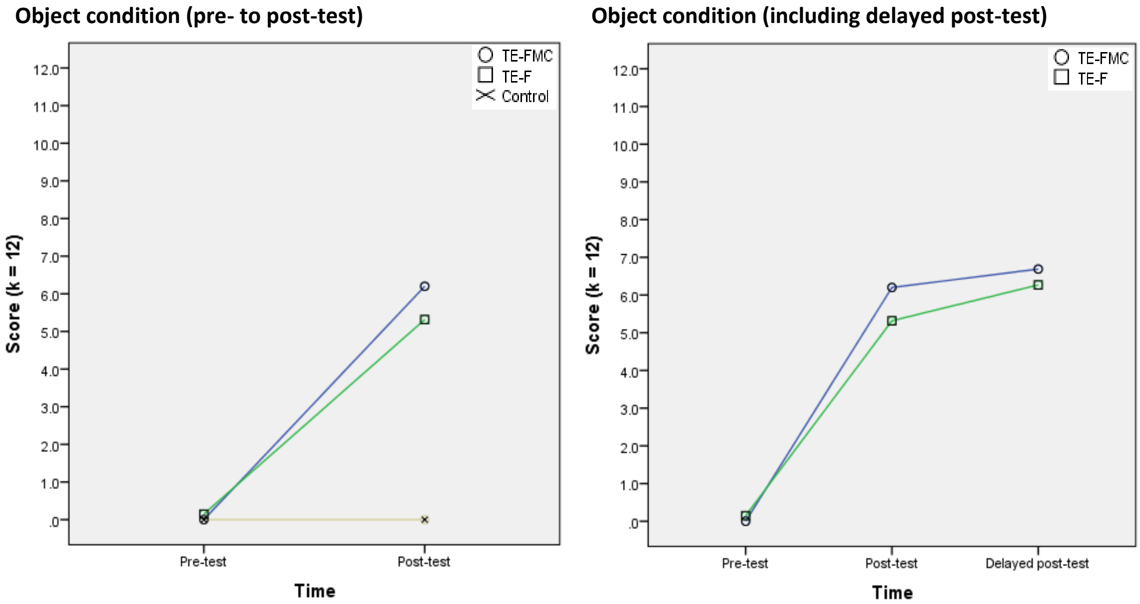


Figure 4.9b: Scores on the Object test condition (Act-Out Production)

Over time (pre-, post- to delayed post-test)

A significant change in scores over pre-, post- and delayed post-test was found for both the TE-FMC (Friedman's ANOVA, χ^2 (2) = 47.328, *p* = .001) and TE-F groups (χ^2 (2) = 36.845, *p* = .001). Pairwise comparisons revealed a significant difference between pre- and post-test (TE-FMC, *z* = -4.796, *p* = .001, *r* = -.51; TE-F, *z* = -4.086, *p* = .001, *r* = -.45) and pre- and delayed post-test (TE-FMC, *z* = -5.007, *p* = .001, *r* = -.53; TE-F, *z* = -4.693, *p* = .001, *r* = -.52) for both groups, and no difference between post- and delayed post-test (TE-FMC, *z* = -0.211, *p* = .833, *r* = -.02; TE-F, *z* = -0.607, *p* = .544, *r* = -.07).

Between groups (pre- and post-test)

All three groups were performing at a comparable level at pre-test (Kruskall Wallis, *H*(2) = 4.47, *p* = .107), with minimal use of ***den*** (Figure 4.9b). There were two instances of a TE-F group participant correctly producing the object article in a sentence at pre-test; however these were isolated cases and there was no evidence of system learning (see section 4.4.2.4). In contrast, a significant difference had developed between the three groups at post-test (*H*(2) = 49.88, *p* = .001), with the Control group scores being significantly lower than both the TE-FMC (*p* = .001, *r* = .67) and TE-F groups' (*p* = .001, *r* = .61) scores.

Between groups (delayed post-test)

No difference was found between the TE-FMC and TE-F groups at delayed post-test (Mann Whitney, *U* = 859.000, *z* = -0.564, *p* = .573, *r* = -.06).

4.4.2.3 Examination of effect sizes

The effect sizes revealed a large effect of the TE-FMC and TE-F interventions over the Control group at post-test (Table 4.9). Both the TE-FMC and TE-F interventions resulted in large changes in the learners' scores between pre- and post-test, as well as pre- and delayed post-test. In contrast the change in the Control group's scores was small. The pre- to delayed post-test effect size calculated for the TE-FMC group was larger than for the TE-F group, reflecting a slight divergence between the two groups' scores at delayed post-test (see Figure 4.8).

Table 4.9: Magnitude of instructional effect on the Act-Out Production task

Contrast	Group(s)	*d*
Between groups	**TE-FMC vs. Control**	1.63
	TE-F vs. Control	1.35
Pre to post	**TE-FMC**	1.62
	TE-F	1.24
	Control	0.05
Pre to delayed post	**TE-FMC**	1.83
	TE-F	1.34

4.4.2.4 Analysis of article use

It was hypothesised that all three groups would be performing at ceiling level on the subject condition (k = 12) at pre-test, given that the article ***der*** was already familiar to the learners from the pre-teaching phase. The descriptive data (Table 4.8, Figure 4.9a), however, suggested that all groups were performing below ceiling level on this test condition at pre-test. In order to explore the nature of the errors the participants were making, the data for each group was examined in order to determine which articles the participants were using when producing their sentences at pre-test and post-test. Table 4.10 presents the percentages[62] of article use (***der***, ***den***, ***other*** or ***missing article***) in the subject and object positions respectively.

Table 4.10: Percentage article use (Act-Out Production)

Group	N	Subject											
		Pre-test				Post-test				Delayed post-test			
		der	den	other	miss	der	den	other	miss	der	den	other	miss
TE-FMC	45	81.1	-	16.5	2.4	91.7	5.7	2.6	-	95.6	1.7	2.7	0.0
TE-F	41	85.8	0.2	13.8	0.2	91.1	4.5	4.4	-	88.8	7.3	3.9	0.0
Control	46	82.2	-	17.8	-	83.3	0.6	14.5	1.6	-	-	-	-
Total	*	82.9	0.1	16.1	0.9	88.6	3.5	7.3	0.6	92.3	4.4	3.3	0.0
Group	**N**	**Object**											
TE-FMC	45	66.1	-	27.0	6.9	46.5	51.5	1.6	0.4	40.2	55.9	3.9	0.0
TE-F	41	73.8	1.2	23.2	1.8	48.0	44.3	7.7	0.0	37.6	52.4	10.0	0.0
Control	46	74.1	-	23.4	2.5	71.2	-	26.6	2.2	-	-	-	-
Total	*	71.2	0.4	24.6	3.8	55.6	31.3	12.2	0.9	39.0	54.3	6.7	0.0

* Pre- and post-test, N = 132; Delayed post-test, N = 86

[62] (Total number of times article used / Total number of articles produced)*100

i) Subject condition

Use of other articles (pre- to post-test)

There was much greater variation in the articles which the participants were using for the subject condition at pre-test; for instance 16.5% of the articles produced by the TE-FMC group at pre-test used an article other than ***der,*** such as ***die*** and ***das*** (feminine and neuter articles in German), ***un*** from French, and ***the*** from English, compared to only 2.4% at post-test. A similar pattern was observed for the TE-F group (pre-test, 13.8%; post-test, 4.4%). This variation accounted for the learners' lower than expected performance on the subject condition at pre-test. Notably, the Control group's use of other articles for the subject condition remained high at post-test (14.5%).

Use of other articles (delayed post-test)

Both the TE-FMC group (95.6%) and TE-F group (88.8%) were reliably employing the article ***der*** for the subject condition at delayed post-test (Table 4.10). There were only a small number of instances of the learners using an alternative article (***die***, ***das***, ***the***) for the subject position (TE-FMC, 2.7%; TE-F, 3.9%).

*Incorrect use of **den** (pre- to post-test)*

At pre-test, no difference was found between the three groups in terms of their incorrect use of ***den*** (Kruskall Wallis, $H(2) = 2.22$, $p = .330$), which was minimal (Table 4.11). At post-test, the difference between the three groups approached significance ($H(2) = 5.924$, $p = .052$). There was a significant increase in both the TE-FMC (Wilcoxon signed rank, $T = 36.00$, $p = .011$, $r = .27$) and TE-F groups' ($T = 21.00$, $p = .027$, $r = .24$) use of ***den*** for the subject, but no change in the Control group's use of ***den***, which remained at 0 ($T = 1.00$, $p = .317$, $r = .0.10$).

Table 4.11: Incorrect use of den for the subject condition (Act-Out Production)

Group	N	Pre-test			Post-test			Delayed post-test		
		M	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	45	0.00	0.00	0	0.69	2.33	0	0.20	0.51	0
TE-F	41	0.02	0.16	0	0.54	1.98	0	0.88	2.69	0
Control	46	0.00	0.00	0	0.07	0.44	0	-	-	-
Total	*	0.01	0.09	0	0.42	1.78	0	0.52	1.91	0

* Pre- and post-test, *N* = 132; Delayed post-test, *N* = 86

*Incorrect use of **den** (pre-, post- to delayed post-test)*

A significant overall change was observed in the TE-FMC group's overuse of ***den*** at the three time points (Friedmans' ANOVA, χ^2 (2) = 8.750, $p = .013$); however pairwise comparisons revealed no differences between any of the time points. Similarly a significant overall change was observed for the TE-F group (Friedmans' ANOVA, χ^2 (2) = 7.000, $p = .030$), but no difference between the three individual time points. Analysis revealed no difference between the two groups (Mann Whitney, U = 929.500, z = .097, p = .923, r = .01). Nevertheless the descriptive statistics (Tables 4.10, 4.11) indicated that there was a tendency towards the TE-F group overusing the object article ***den***, to a greater extent than the TE-FMC group at delayed post-test. There was an increase in the TE-FMC group's overuse of ***den*** at post-test (0% to 4.7%), however this had decreased at delayed post-test (to 1.7%). In contrast the TE-F group's use of ***den*** for the subject condition increased both from pre- to post-test (0.2% to 4.5%), and post- to delayed post-test (to 7.3%). These results further support the finding (section 4.4.2.2) that the TE-FMC group's accuracy on the subject condition had improved at delayed post-test.

ii) Object condition

Use of other articles (pre- to post-test)

At pre-test there were frequent instances of 'other' articles (e.g. ***die***, ***das***, ***the***) being used in the object condition (TE-FMC, 27.0%; TE-F, 23.2%; Control, 23.4%). At post-test, however, the TE-FMC (1.6%) and TE-F groups' (7.7%) use of 'other' articles had decreased. Further there was a substantial increase in the correct use of the object article ***den*** by the TE-FMC (51.5%) and TE-F groups (44.3%), compared to zero instances by the Control group. This increased accuracy was reflected in the significant improvement made by the TE-FMC and TE-F groups in their production of ***den*** in the object condition at post-test (section 4.4.2.2).

Use of other articles (delayed post-test)

At delayed post-test the incorrect use of 'other' articles for the object test items was low in the TE-FMC group (3.9%), although slightly higher for the TE-F group (10%). Both the TE-FMC (55.9%) and TE-F groups (52.4%) continued to correctly employ ***den*** in a majority of obligatory cases.

*Incorrect use of **der** (pre- to post-test)*

At pre-test there was a strong tendency across all the three groups to use the subject article ***der*** for the object of the sentence (TE-FMC, 66.1%; TE-F, 73.8%; Control group, 74.1%) (Table 4.10, above). Despite the significant improvement in the TE-FMC and TE-F groups' use of the object article at post-test (section 4.4.2.2), the instances of the subject article ***der*** being used for the object position remained high at post-test (TE-FMC, 46.5%; TE-F, 48.0%).

*Incorrect use of **der** (delayed post-test)*

At delayed post-test, as at post-test, there was a relatively high percentage of incorrect use of the subject article (***der***) for the object condition (TE-FMC, 40.2%; TE-F group in 37.6%) (Table 4.10). However it is important to note that this finding (both at post- and delayed post-test) was not due to there being only chance level accuracy in the learners' use of the newly learned object article ***den***. Rather it was found that at post-test approximately half of the participants in the TE-FMC and TE-F groups were consistently using ***den*** correctly, whereas the remaining participants continued to use the subject article ***der*** or an alternative (***die***, ***das***, ***the***) article. This finding is explored in more detail in section 5.2.

4.4.3 *Sentence Repetition* task

4.4.3.1 Analysis of total scores

The descriptive statistics relating to the learners' overall performance on the Sentence Repetition task are detailed in Table 4.12 (see also Figure 4.10):

Table 4.12: Descriptive statistics for the Sentence Repetition task

Group	N	Total (*k* = 12)								
		Pre-test			Post-test			Delayed post-test		
		M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
TE-FMC	45	5.67	2.2	6	8.8	2.19	9	9.13	2.28	10
TE-F	41	5.81	1.25	6	8	2.2	8	8.22	2.20	8
Control	46	5.37	1.55	6	5.67	1.63	6	-	-	-
Total	*	5.61	1.72	6	7.46	2.42	7	8.70	2.28	9
Group	**N**	**SVO+Subj (*k* = 3)**								
TE-FMC	45	2.6	0.65	3	2.47	0.73	3	2.44	0.91	3
TE-F	41	2.76	0.49	3	2.54	0.9	3	2.10	1.16	3
Control	46	2.3	0.94	3	2.54	0.75	3	-	-	-
Total	*	2.55	0.75	3	2.52	0.79	3	2.28	1.05	3
Group	**N**	**OVS+Subj (*k* = 3)**								
TE-FMC	45	1.98	1.1	2	2.47	0.87	3	2.69	0.56	3
TE-F	41	2.29	0.9	3	2.32	0.99	3	2.39	0.86	3
Control	46	2.01	1.05	2	2.28	0.96	3	-	-	-
Total	*	2.11	1.02	2	2.36	0.93	3	2.55	0.73	3
Group	**N**	**SVO+Obj (*k* = 3)**								
TE-FMC	45	0.33	0.6	0	1.73	1.14	2	1.69	1.28	2
TE-F	41	0.32	0.65	0	1.37	1.07	1	1.59	1.14	2
Control	46	0.22	0.59	0	0.11	0.38	0	-	-	-
Total	*	0.29	0.61	0	1.05	1.15	1	1.64	1.21	2
Group	**N**	**OVS+Obj (*k* = 3)**								
TE-FMC	45	0.76	1.05	0	2.13	1.1	2	2.31	1.12	3
TE-F	41	0.44	0.63	0	1.78	1.21	2	2.15	1.22	3
Control	46	0.76	0.87	1	0.74	1.02	0	-	-	-
Total	*	0.66	0.88	0	1.54	1.26	2	2.23	1.16	3

* Pre- and post-test, *N* = 132; Delayed post-test, *N* = 86

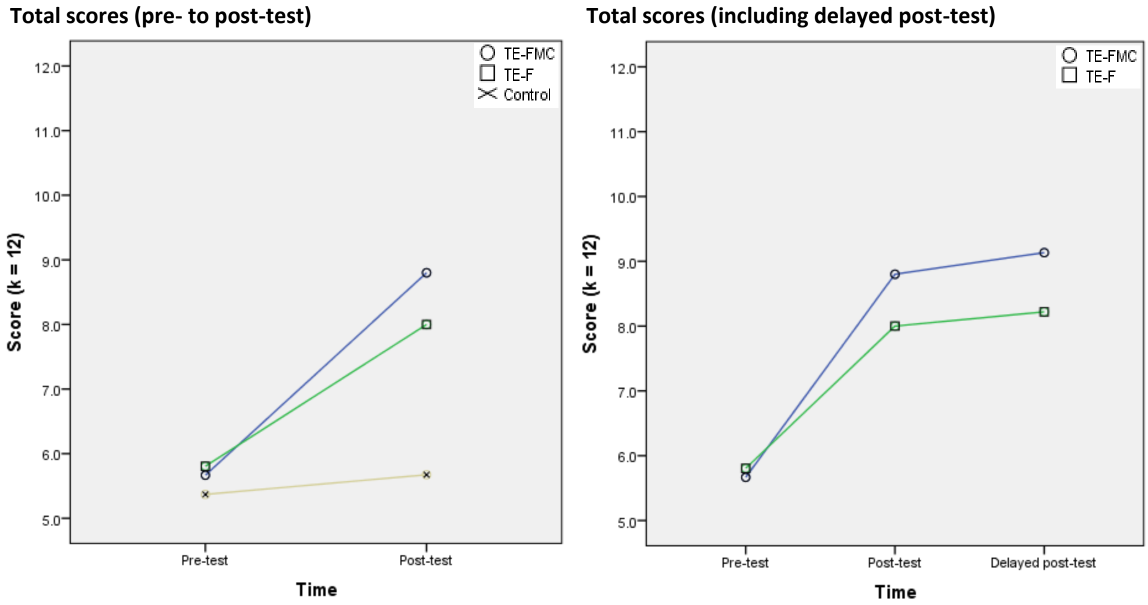


Figure 4.10: Total scores on the Sentence Repetition task

Over time (pre- to post-test)

Analysis revealed a significant improvement in the learners' performance between pre- and post-test for both the TE-FMC (Wilcoxon signed rank, T = 897.00, p = .001, r = .59) and the TE-F group (T = 615.50, p = .001, r = .55). In contrast, no change over time was found for the Control group (T = 320.00, p = .284, r = .11).

Over time (pre-, post- to delayed post-test)

There was a significant change in the TE-FMC group's performance on this task (Friedman's ANOVA, χ^2 (2) = 57.268, p = .001). Pairwise comparisons revealed a significant improvement between pre- and post-test (z = -5.639, p = .001, r = -.59) and pre- and delayed post-test (z = -6.377, p = .001, r = -.67), but no change between post- and delayed post-test (z = -0.738, p = .461, r = -.08). A similar pattern was observed in the TE-F group. There was a significant change in their scores over the three time points (χ^2 (2) = 50.936, p = .001), with a significant improvement between pre- and post-test (z = -5.466, p = .001, r = -.60) as well as delayed post-test (z = -5.963, p = .001, r = -.66), but no difference between post- and delayed post-test (z = -0.497, p = .619, r = -.05).

Between groups (pre- and post-test)

Analysis confirmed that at pre-test the three groups were performing at an equivalent level (Kruskall Wallis, H(2) = 1.42, p = .493). By post-test a significant difference had developed between the groups (H(2) = 44.30, p = .001). Pairwise comparisons revealed that the TE-FMC (p = .001, r = .66) and TE-F groups (p = .001, r = .53) were outperforming the Control group.

Between groups (delayed post-test)

An examination of the descriptive statistics (Table 4.12, Figure 4.10) suggested that the TE-FMC group was marginally outperforming the TE-F group at delayed post-test. In line with this observation, the Mann Whitney U test was approaching significance (U = 720.500, z = -1.768, p = .077, r = -.19), suggesting that although both groups maintained the gains at delayed post-test, the TE-FMC group had made slightly larger gains.

4.4.3.2 Analysis by test condition (SVO+Subj, OVS+Subj, SVO+Obj, OVS+Obj)

The Sentence Repetition task contained test sentences in both SVO and OVS word orders (k = 3 for each condition). Participants were marked on their production of both ***der*** and ***den*** within each sentence resulting in four test conditions; SVO+Subj, OVS+Subj, SVO+Obj, OVS+Obj (k = 3 for each). The descriptive statistics for each test condition are detailed in Table 4.10 (above).

i) SVO+Subj condition

Over time (pre- to post-test)

No change over time was found for the three groups between pre- and post-test (TE-FMC, Wilcoxon signed rank, T = 69.50, p = .280, r = -.11; TE-F, T = 29.00, p = .130, r = -.17; Control group, T = 119.50, p = .121, r = .16) (see also Figure 4.11a).

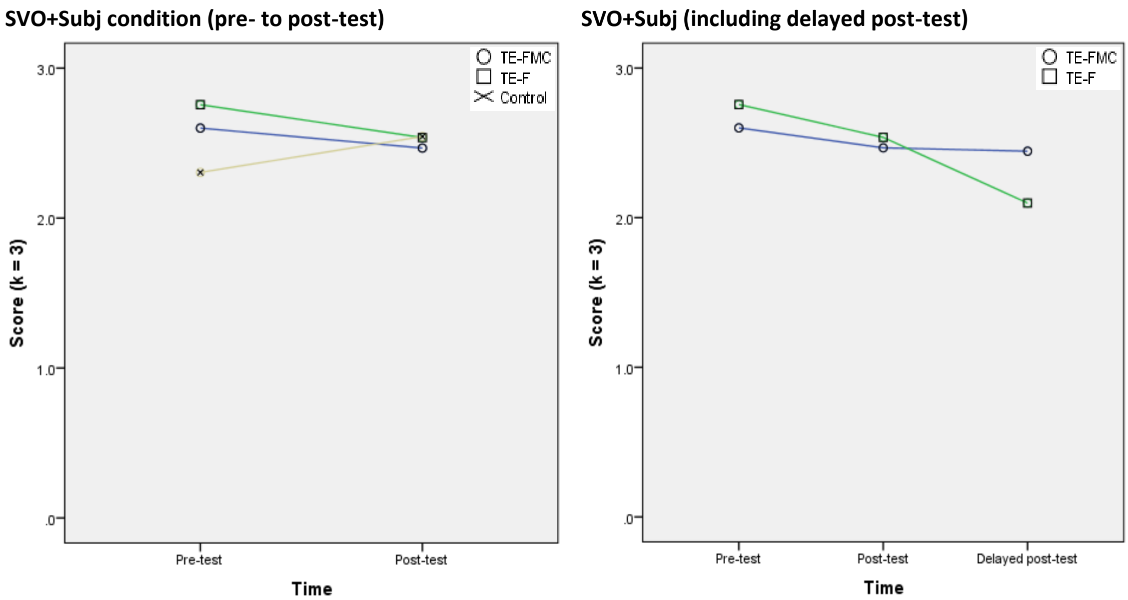


Figure 4.11a: Scores on the SVO+Subj test condition (Sentence Repetition)

Over time (pre-, post- to delayed post-test)

The TE-F group's scores decreased significantly over the three time points (Friedman's ANOVA, χ^2 (2) = 10.603, p = .005) (Table 4.12, Figure 4.11a). Pairwise comparisons, however, found no difference between the individual time points. In contrast, the TE-FMC group maintained their ceiling level performance on this condition across the three time points (χ^2 (2) = 0.804, p = .669).

Between groups (pre-to post-test)

Analysis revealed a significant difference between the three groups at pre-test (Kruskall Wallis, $H(2)$ = 6.60, p = .037); an examination of the descriptive statistics (Table 4.12, Figure 4.11a) found that the Control group's scores were significantly lower than the TE-F group's scores (p = .011, r = .27). At post-test no difference was found between the three groups ($H(2)$ = 1.18, p = . 554).

Between groups (delayed post-test)

Despite the decrease in the TE-F group's scores on this condition, no difference was found between the TE-FMC and TE-F groups at delayed post-test (Mann Whitney, U = 778.500, z = -1.419, p = .156, r = -.15).

ii) OVS+Subj condition

Over time (pre- to post-test)

The TE-FMC group's scores significantly improved between pre- and post-test (Wilcoxon signed rank, T = 317.00, p = .007, r = .28). In contrast no significant change over time was observed for either the TE-F group (T = 142.50, p = .886, r = .02) or the Control group (T = 142.50, p = .140, r = .15). The descriptive data (Table 4.12, Figure 4.11b) revealed that all three groups were performing at below ceiling level on this condition at pre-test, which may have been due to the subject article occurring in the less salient sentence medial position within the OVS test items.

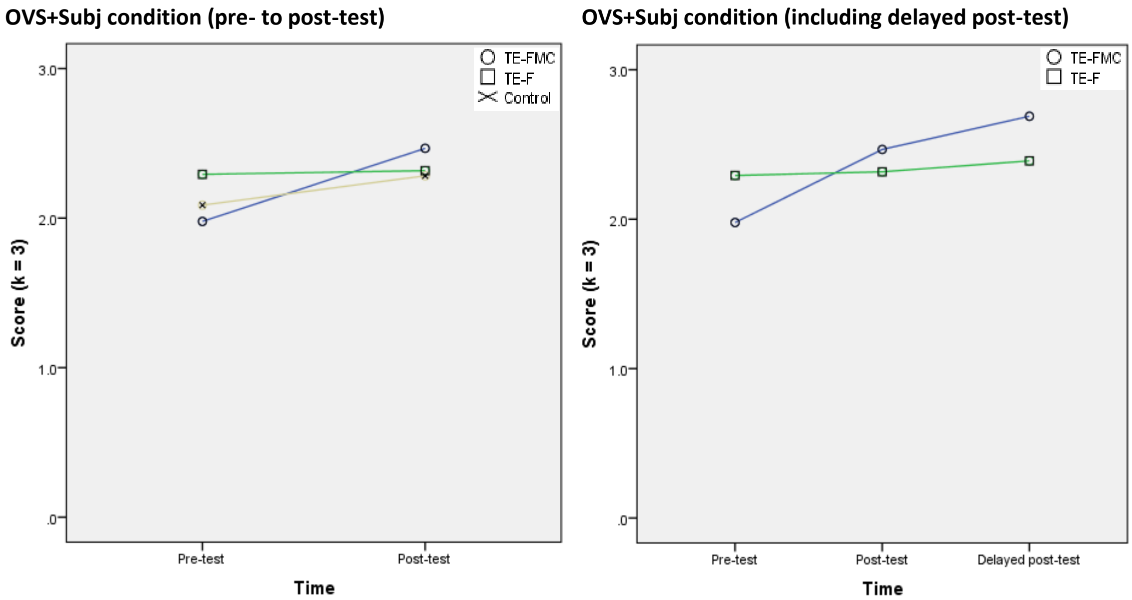


Figure 4.11b: Scores on the OVS+Subj test condition (Sentence Repetition)

Over time (pre-, post- to delayed post-test)

The TE-FMC group performance on the OVS+Subj condition had improved over the three time points (Friedman's ANOVA, χ^2 (2) = 14.000, p = .001); a significant difference was found between the TE-FMC group's pre- and delayed post-test performance on this test condition (z = -2.846, p = .004, r = -.30). For the TE-F group, however, there was no change in their scores over time (χ^2 (2) = 0.022, p = .989) (Figure 4.11b).

Between groups (pre-to post-test)

Despite the significant improvement observed over in the TE-FMC group's performance over time, the scores of the three groups were found to be equivalent at both pre- (Kruskall Wallis, H(2) = 1.72, p = .424) and post-test (H(2) = 1.08, p = .583).

Between groups (delayed post-test)

There was no difference between the TE-FMC and TE-F groups' scores at delayed post-test (Mann Whitney, U = 765.000, z = -1.635, p = .102, r = -.18).

iii) SVO+Obj condition

Over time (pre- to post-test)

Tthe TE-FMC (Wilcoxon signed rank, T = 561.00, p = .001, r = .54) and TE-F groups' (T = 419.00, p = .001, r = .49) scores increased significantly at post-test, whereas there was no significant change in the Control group's scores (T = 23.00, p = .353, r = -.10) (Table 4.12, Figure 4.11c).

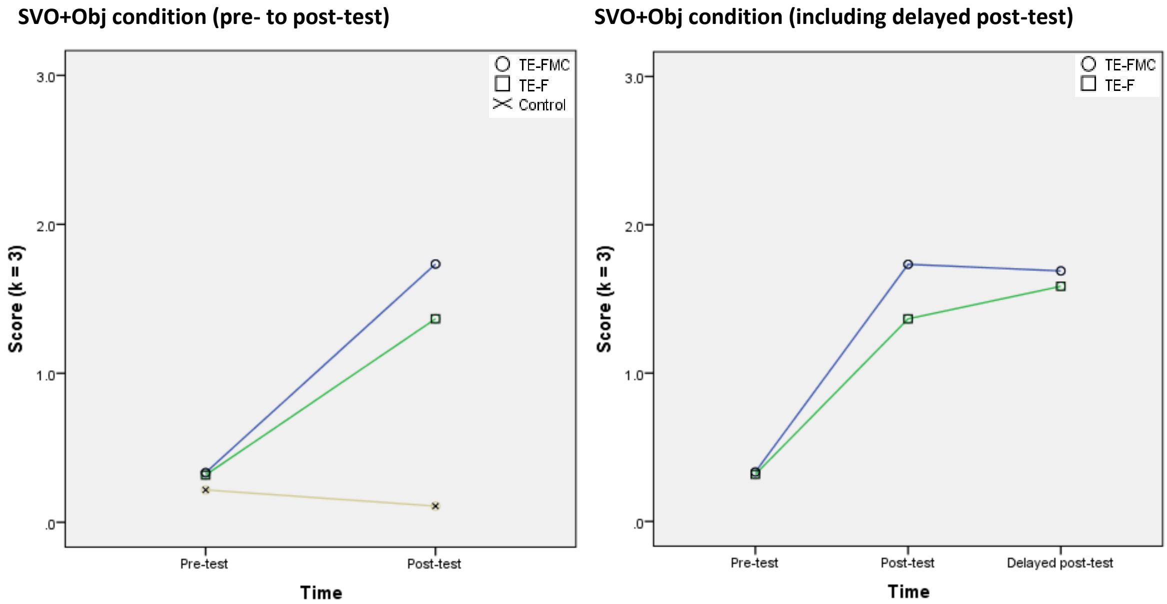


Figure 4.11c: Scores on the SVO+Obj test condition (Sentence Repetition)

Over time (pre-, post- to delayed post-test)

In terms of their performance on the SVO+Obj condition over the three time points (Figure 4.11c), analysis revealed a significant change for both the TE-FMC and TE-F groups (TE-FMC, Friedman's ANOVA, χ^2 (2) = 48.439, *p* = .001; TE-F, χ^2 (2) = 35.504, *p* = .001). A significant difference was found between both group's pre- and post-test (TE-FMC, *z* = -5.060, *p* = .001, *r* = -.53; TE-F, *z* = -3.920, *p* = .001, *r* = -.43) and pre- and delayed post-test scores (TE-FMC, *z* = -4.902, *p* = .001, *r* = -.52; TE-F, *z* = -4.693, *p* = .001, *r* = -.52) and no difference between post- and delayed post-test (TE-FMC, *z* = .158, *p* = .874, *r* = .02; TE-F, *z* = -0.773, *p* = .440, *r* = -.09).

Between groups (pre-to post-test)

There was no difference between the three groups at pre-test (Kruskall Wallis, *H*(2) = 1.81, *p* = .405); however at post-test a significant main effect of Group was found (*H*(2) = 52.49, *p* = .001). Pairwise comparisons revealed that at post-test the Control group was performing at a significantly lower level than both the TE-FMC (*p* = .001, *r* = .72) and TE-F groups (*p* = .001, *r* = .58).

Between groups (delayed post-test)

There was no difference between the TE-FMC and TE-F groups' scores at delayed post-test (Mann Whitney, *U* = 865.00, *z* = -0.517, *p* = .605, *r* = -.06).

iv) OVS+Obj condition

Over time (pre- to post-test)

A significant improvement over time was found for the TE-FMC (Wilcoxon signed rank, *T* = 549.00, *p* = .001, *r* = .51) and TE-F groups (*T* = 483.00, *p* = .001, *r* = .52) between pre- and post-test. In contrast, there was no change in the Control group scores (*T* = 197.00, *p* = .881, *r* = -.02) (Table 4.12 above, Figure 4.11d).

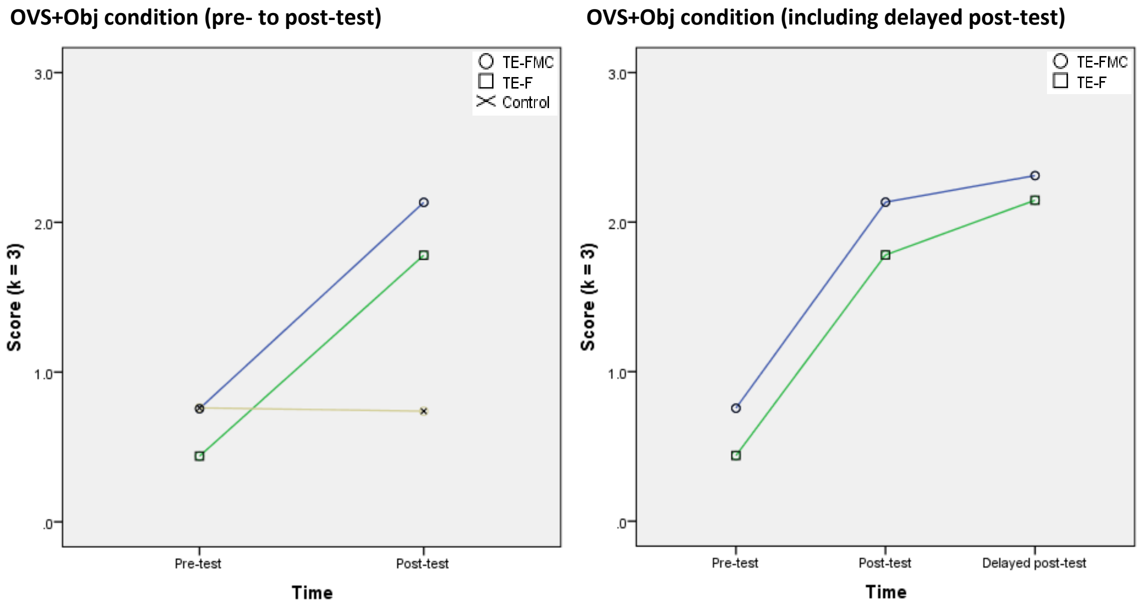


Figure 4.11d: Scores on the OVS+Obj test condition (Sentence Repetition)

Over time (pre-, post- to delayed post-test)

Both the TE-FMC (Friedman's ANOVA, χ^2 (2) = 41.392, *p* = .001) and TE-F groups' scores (χ^2 (2) = 48.123, *p* = .001) improved significantly over the three time points (Figure 4.11d). A significant difference was found between pre- and post-test (TE-FMC, *z* = -4.164, *p* = .001, *r* = -.44; TE-F, *z* = -4.252, *p* = .001, *r* = -.47) and pre- and delayed post-test for both groups (TE-FMC, *z* = -5.007, *p* = .001, *r* = -.53; TE-F, *z* = -5.522, *p* = .001, *r* = -.61); however there was no change in either groups' performance between post- and delayed post-test (TE-FMC, *z* = -0.843, *p* = .399, *r* = -.09; TE-F, *z* = -1.270, *p* = .204, *r* = -.14).

Between groups (pre-to post-test)

No difference was found between the three groups at pre-test (Kruskall Wallis, *H*(2) = 2.87, *p* = .238), whereas a significant difference was observed at post-test (*H*(2) = 29.32, *p* = .001). Pairwise comparisons yielded a significant difference between the Control group and both the TE-FMC (*p* = .001, *r* = .55) and TE-F groups (*p* = .001, *r* = .41).

Between groups (delayed post-test)

There was no difference between the TE-FMC and TE-F groups' performances on the OVS+Obj condition at delayed post-test (Mann Whitney, *U* = 849.500, *z* = -0.745, *p* = .456, *r* = -.08).

4.4.3.3 Examination of effect sizes

Both the TE-FMC and TE-F interventions resulted in a large effect in comparison to the Control group at post-test (Table 4.13). This effect was also reflected in the large magnitude of change for both the TE-FMC and TE-F groups between pre- and post-test as well as between pre- and delayed post-test. In contrast the magnitude of change for the Control group was small.

Table 4.13: Magnitude of instructional effect on the Sentence Repetition task

Contrast	Group(s)	*d*
Between groups	**TE-FMC vs. Control**	1.64
	TE-F vs. Control	1.23
Pre to post	**TE-FMC**	1.45
	TE-F	1.27
	Control	0.19
Pre to delayed post	**TE-FMC**	1.54
	TE-F	1.40

4.4.3.4 Analysis of article use

The learners' article use on the Sentence Repetition task was examined, in order to explore the nature of the errors the participants were making on the respective test conditions at each time point. Table 4.14 details the overall frequency (%)[63] with which each article (***der***, ***den***, ***other*** or ***missing article***) was used by the three groups in the respective test conditions (SVO+Subj, SVO+Obj, OVS+Subj, OVS+Obj).

Table 4.14: Percentage article use (Sentence Repetition)

Group	N	Pre-test der	Pre-test den	Pre-test other	Pre-test miss	Post-test der	Post-test den	Post-test other	Post-test miss	Delayed post-test der	Delayed post-test den	Delayed post-test other	Delayed post-test miss
						SVO+Subj							
TE-FMC	45	86.7	3.7	9.6	0.0	80.7	18.5	0.7	0.0	80.7	18.5	0.8	0.0
TE-F	41	93.5	3.3	3.3	0.0	82.9	13.0	3.3	0.8	70.7	28.5	0.0	0.8
Control	46	76.8	10.9	10.1	2.2	84.8	7.2	6.5	6.5	-	-	-	-
Total	*	85.4	6.1	7.8	0.8	82.8	12.9	3.5	0.8	76.0	23.3	0.4	0.4
Group	**N**					**OVS+Subj**							
TE-FMC	45	67.4	0.7	20.7	11.1	81.5	11.1	4.4	3.0	87.4	8.1	2.3	2.2
TE-F	41	78.9	0.8	13.8	6.5	77.2	17.9	4.1	0.8	79.7	15.4	4.9	0.0
Control	46	69.6	2.2	16.7	11.6	76.8	2.2	18.8	2.2	-	-	-	-
Total	*	71.7	1.3	17.2	9.8	78.5	10.1	9.3	2.0	83.7	11.6	3.5	1.2
Group	**N**					**SVO+Obj**							
TE-FMC	45	53.3	11.9	26.7	8.1	37.8	56.3	4.4	1.5	41.5	55.6	3.0	0.0
TE-F	41	66.7	8.9	17.1	7.3	44.7	45.5	6.5	3.3	44.7	52.0	3.3	0.0
Control	46	58.7	6.5	28.3	6.5	68.8	3.6	26.1	1.4	-	-	-	-
Total	*	59.3	9.1	24.2	7.3	50.8	34.6	12.6	2.0	43.0	53.9	3.1	0.0
Group	**N**					**OVS+Obj**							
TE-FMC	45	67.4	25.2	5.9	1.5	25.2	69.6	3.0	2.2	23.0	74.8	0.0	2.2
TE-F	41	77.2	13.8	8.9	0.0	38.2	59.3	2.4	0.0	26.8	71.5	1.6	0.0
Control	46	61.6	25.4	10.9	2.2	65.2	25.4	7.2	2.2	-	-	-	-
Total	*	68.4	21.7	8.6	1.3	43.2	51.0	4.3	1.5	24.8	73.3	0.8	1.2

* Pre- and post-test, *N* = 132; Delayed post-test, *N* = 86

i) SVO+Subj condition

Use of other articles (pre- to post-test)

There was a higher percentage of instances of the Control group producing 'other' articles for the SVO+Subj condition at both pre- (10.1%) and post-test (6.5%), than for either the TE-FMC or TE-F groups. Further the Control group correctly produced the subject article ***der*** in only 76.8% of cases (TE-FMC, 86.7%; TE-F, 93.5%), reflecting the significant difference observed between groups at pre-test (see section 4.4.3.2).

Use of other articles (delayed post-test)

There was little or no use of 'other' articles (***die***, ***das***, ***the***) by either the TE-FMC or TE-F group at delayed post-test (Table 4.14).

[63] (Total number of times article used / Total number of articles produced)*100

*Incorrect use of **den** (pre- to post-test)*

There was substantially more incorrect use of the object article ***den*** by the Control group (10.9%) compared to the TE-FMC (3.7%) and TE-F groups (3.3%) at pre-test. This difference was approaching significance (Kruskall Wallis, $H(2) = 5.267$, $p = .072$). There was also a significant difference between the groups at post-test ($H(2) = 6.527$, $p = .038$); however, this was due to the TE-FMC group making significantly more (incorrect) use of the article ***den*** than the Control group ($p = .013$, $r = .26$). Further there was a significant increase in both the TE-FMC (Wilcoxon signed rank, $T = 170.50$, $p = .002$, $r = .33$) and TE-F groups' ($T = 66.00$, $p = .030$, $r = .24$) incorrect use of ***den*** for this condition at post-test (Table 4.15).

Table 4.15: Incorrect use of den for the SVO+Subj and OVS+Subj conditions (Sentence Repetition)

Group	N	Pre-test			Post-test			Delayed post-test		
		SVO+Subj (*k* = 3)								
		M	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	45	0.11	0.38	0	0.56	0.76	0	0.56	0.92	0
TE-F	41	0.10	0.30	0	0.39	0.83	0	0.85	1.13	0
Control	46	0.33	0.70	0	0.22	0.55	0	-	-	-
Total	*	0.18	0.51	0	0.39	0.73	0	0.7	1.03	0
Group	**N**	**OVS+Subj (*k* = 3)**								
TE-FMC	45	0.02	0.15	0	0.33	0.77	0	0.24	0.53	0
TE-F	41	0.02	0.16	0	0.54	0.87	0	0.46	0.75	0
Control	46	0.07	0.44	0	0.07	0.25	0	-	-	-
Total	*	0.04	0.29	0	0.30	0.70	0	0.35	0.65	0

* Pre- and post-test, $N = 132$; Delayed post-test, $N = 86$

*Incorrect use of **den** (pre-, post- to delayed post-test)*

Analysis revealed a significant increase for both groups in their overuse of ***den*** over the three time points (TE-FMC, $\chi^2(2) = 11.488$, $p = .003$; TE-F, $\chi^2(2) = 15.559$, $p = .001$); although pairwise comparisons revealed no differences between the individual time points for the TE-FMC group. In contrast the TE-F group's incorrect use of ***den*** significantly increased between pre- and delayed post-test ($z = -2.540$, $p = .011$, $r = -.28$). Indeed, the TE-F group were incorrectly using the object article ***den*** to a greater extent (28.5%) than the TE-FMC group (18.5%) on this condition at delayed post-test.

ii) OVS+Subj condition

Use of other articles (pre- to post-test)

At pre-test the correct use of the subject article ***der*** on the OVS+Subj condition was notably low (TE-FMC, 67.4%; TE-F, 78.9%; Control, 69.6%) and there was a high frequency of 'other' articles being used (TE-FMC, 20.7%; TE-F, 13.8%; Control, 16.7%). This finding may have been due to the fact that in the OVS sentences ***der*** appeared in sentence medial position and was therefore less salient. At post-test, whilst the Control group's use of 'other' articles remained high (18.5%), the TE-FMC and TE-F groups' usage had decreased substantially.

Use of other articles (delayed post-test)

At delayed post-test, the TE-FMC and TE-F groups' use of 'other' articles remained low (Table 4.14 above).

*Incorrect use of **den** (pre- to post-test)*

A significant increase in the incorrect use of ***den*** was found for both the TE-FMC (Wilcoxon signed rank, T = 51.50, p = .012, r = .26), and TE-F groups (T = 105.00, p = .001, r = .38), whereas there was no change in the Control group's infrequent use of ***den*** at post-test (T = 6.00, p = .705, r = .04) (Table 4.14, Table 4.15). Whilst there was no difference between the three groups at pre-test (Kruskall Wallis, $H(2)$ = 0.006, p = .997), a significant difference emerged at post-test ($H(2)$ = 10.809, p = .004), with the TE-F group's overuse of the object article ***den*** being significantly higher than that of the Control group (p = .001, r = .35).

*Incorrect use of **den** (delayed post-test)*

The TE-FMC group's use of ***der*** was more accurate at delayed post-test (87.4%) than at pre-test (67.4%), reflecting the TE-FMC group's significant improvement on this condition over the three time points (see section 4.4.3.2). Nevertheless, there was a significant increase in both the TE-FMC (Friedman's ANOVA, χ^2 (2) = 7.423, p = .024) and TE-F groups' (χ^2 (2) = 18.542, p = .001) incorrect use of ***den*** (Table 4.15). For the TE-FMC group, pairwise comparisons revealed no differences between the individual time points. For the TE-F group, however, the difference was approaching significance between the pre- and post-test (z = -2.264, p = .001, r = -25) and pre- and delayed post-test (z = -2.209, p = .027, r = -.24). ***den*** was overused to a greater extent by the TE-F group (15.4%) than the TE-FMC group (8.1%) at delayed post-test, although this difference was not significant (Mann Whitney, U = 1044.000, z = 1.206, *ns*, r = .13).

iii) SVO+Obj condition

Use of other articles (pre- to post-test)

The correct production of ***den*** was low for all three groups at pre-test and there was a high percentage use of 'other' articles (Table 4.14 above). By post-test the TE-FMC and TE-F groups' use of 'other' articles had decreased but remained high for the Control group (Table 4.14). This finding reflected the significant improvement in the TE-FMC and TE-F groups' (but not the Control group's) correct use of ***den*** at post-test (section 4.4.3.2).

Use of other articles (delayed post-test)

As at post-test, the TE-FMC and TE-F groups' use of 'other' articles remained low at delayed post-test (Table 4.14).

Incorrect use of ***der*** *(pre- to post-test)*

At pre-test, a majority of the errors made by all three groups were due to overuse of ***der*** (Table 4.14 above). At post-test, the Control group continued to rely primarily on ***der*** for the SVO+Obj condition, whereas the TE-FMC and TE-F groups' use of ***der*** had decreased to a certain extent, in line with the significant improvement in their performance on this condition (section 4.4.3.2).

Incorrect use of ***der*** *(delayed post-test)*

At delayed post-test, both the TE-FMC and TE-F groups were correctly supplying ***den*** for a majority of SVO+Obj test items. Nevertheless, the instances of learners' incorrectly producing ***der*** remained relatively high (Table 4.14).

iv) OVS+Obj condition

Use of other articles (pre- to post-test)

The use of 'other' articles in the OVS+Obj condition was low for all three groups at both pre- and post-test (Table 4.14 above). Notably, for this condition, in which the unfamiliar object article ***den*** was in the more salient sentence initial position, the percentage of correct reproductions of ***den*** at pre-test was higher than anticipated (TE-FMC, 25.2%; TE-F, 13.8%; Control, 25.4%).

Use of other articles (delayed post-test)

There were no instances of the TE-FMC learners producing 'other' articles for this condition at delayed post-test and only 1.6% of cases for the TE-F group.

Incorrect use of ***der*** *(pre- to post-test)*

All three groups used ***der*** for a majority of OVS+Obj test items at pre-test (Table 4.14). Overuse of ***der*** remained high at post-test for the Control group, however had reduced for both the TE-FMC (25.2%) and TE-F (38.2%) groups. In contrast, the percentage correct use of ***den*** by the TE-FMC (69.6%) and TE-F (59.3%) learners increased substantially at post-test, reflecting the learners' significant improvement on this condition (section 4.4.3.2). The Control group's correct use of ***den*** remained much lower at post-test and equivalent to their pre-test performance (25.4%).

*Incorrect use of **der** (delayed post-test)*

At delayed post-test, the TE-FMC and TE-F groups' incorrect use of ***der*** remained relatively low (23.0% and 26.8% respectively), reflecting the learners' sustained improved performance (section 4.4.3.2).

4.4.4 Summary of findings for the oral outcome measures

The analysis of the learners' performance on the time and communicatively pressured oral outcome measures revealed a similar pattern to that of the untimed written tasks. Firstly both the TE-FMC and TE-F groups' overall scores on the respective tasks significantly improved at post-test. Secondly both groups had sustained this improvement at delayed post-test. In contrast, no improvement was observed in the Control group's scores between pre and post-test for any of the oral tasks.

For the Act-Out Comprehension task, the analysis by test condition indicated that the learners' improvement could be attributed to an improvement in their comprehension of the target feature within OVS test items. On the Act-Out Production task, both the TE-FMC and TE-F groups significantly improved in their production of ***den*** at post-test, whereas the Control group performance remained at baseline. Notably the TE-FMC learners' accuracy when producing ***der*** also improved to a certain extent at post-test, as reflected by the inferential statistics and examination of the learners' article use. With regards to the Sentence Repetition task, the TE-FMC and TE-F groups made significant gains in their production of ***den*** in both SVO and OVS test items at post-test. An examination of the learners' article use revealed that there were instances of both groups overgeneralising their use of the target feature to the subject test items at post-test. At delayed post-test, the level of overgeneralisation by the TE-FMC group had decreased, whereas the TE-F learners continued to overuse the target feature to the same extent as at post-test.

4.5 Analysis of Animacy conditions (A+A, A+I, I+A)

Two hypotheses were generated with respect to the learners' use of animacy when interpreting test items on the respective outcome measures (section 3.6.2). If the learners were relying on animacy as well as word order then:

A) At pre-test the learners will have been able to correctly interpret OVS test items containing an Animate subject and Inanimate object
B) The largest gains between pre- and post-test will have been made on test items containing an Inanimate subject and Animate object, across both word order conditions

The analysis presented below investigated hypotheses A and B for the learners' performance on the oral Act-Out Comprehension task. The learners' scores (total and by test condition) on each of the Animacy conditions was analysed for pre- and post-test separately (hypothesis A). Additionally the gains scores (post-test score minus pre-test score) for each condition was analysed (hypothesis B). The same pattern of results was found across all of the outcome measures. Therefore, the analysis of Animacy conditions for the Sentence Matching, Gap-fill and Act-Out Production tasks will not be presented (descriptive statistics can be found in Appendix 24.

4.5.1 Performance on Animacy conditions (A+A, A+I, I+A) at pre- and post-test

4.5.1.1 Analysis of total scores

Table 4.16 presents the descriptive statistics for each of the Animacy conditions at pre- and post-test.

Table 4.16: Descriptive statistics for the Animacy conditions (Act-Out Comprehension)

Group	N	A+A (*k* = 6)					
		Pre-test			Post-test		
		M	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	45	2.91	0.29	3	4.22	1.48	4
TE-F	41	3.02	0.42	3	3.98	1.13	4
Control	46	2.91	0.46	3	2.96	0.42	3
Total	132	2.94	0.40	3	3.70	1.22	3
Group	**N**	**A+I (*k* = 6)**					
TE-FMC	45	3.29	0.51	3	4.18	1.27	4
TE-F	41	3.22	0.69	3	4.32	1.29	4
Control	46	2.83	0.64	3	3.07	0.49	3
Total	132	3.10	0.65	3	3.83	1.21	3
Group	**N**	**I+A (*k* = 6)**					
TE-FMC	45	3.09	0.56	3	4.27	1.29	4
TE-F	41	2.95	0.67	3	4.12	1.23	4
Control	46	2.91	0.55	3	3.15	0.47	3
Total	132	2.99	0.59	3	3.83	1.16	3

Comparing Animacy conditions (pre-test)

No difference was found in either the TE-F (Friedman's ANOVA, $\chi^2(2) = 2.735$, $p = .225$) or Control groups' ($\chi^2(2) = 0.767$, $p = .681$) performance across the conditions at pre-test. For the TE-FMC group, however, a significant difference was observed ($\chi^2(2) = 13.138$, $p = .001$). Nevertheless pairwise comparisons found no difference between any of the individual conditions. The descriptive statistics (Table 4.16) suggested that the overall significant difference was due to the TE-FMC group's lower performance on the A+A condition at pre-test.

Comparing Animacy conditions (post-test)

At post-test, no difference was found between the three Animacy conditions for any of the groups (TE-FMC, Friedman's ANOVA, $\chi^2(2) = 1.641$, $p = .440$; TE-F, $\chi^2(2) = 4.234$, $p = .120$; Control, $\chi^2(2) = 4.269$, $p = .118$).

4.5.1.2 Analysis by test condition (SVO, OVS)

i) SVO condition

Tables 4.17 details each group's scores on the Animacy conditions for the SVO test items.

Table 4.17: Descriptive statistics for the Animacy conditions on the SVO test condition (Act-Out Comprehension)

Group	N	A+A (*k* = 3)					
		Pre-test			Post-test		
		M	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	45	2.91	0.29	3	2.89	0.32	3
TE-F	41	2.89	0.38	3	2.88	0.33	3
Control	46	2.78	0.47	3	2.89	0.31	3
Total	132	2.86	0.39	3	2.89	0.32	3
Group	**N**	**A+I (*k* = 3)**					
TE-FMC	45	3.00	0.00	3	2.93	0.25	3
TE-F	41	2.92	0.26	3	2.90	0.30	3
Control	46	2.74	0.57	3	2.93	0.25	3
Total	132	2.89	0.38	3	2.92	0.27	3
Group	**N**	**I+A (*k*= 3)**					
TE-FMC	45	2.89	0.32	3	2.91	0.29	3
TE-F	41	2.73	0.50	3	2.85	0.36	3
Control	46	2.83	0.44	3	2.96	0.21	3
Total	132	2.82	0.42	3	2.91	2.89	3

Comparing Animacy conditions (pre-test)

Friedman's ANOVA revealed no differences between any of the Animacy conditions for the TE-FMC ($\chi^2(2) = 4.667$, $p = .097$) and Control groups ($\chi^2(2) = 0.340$, $p = .844$) at pre-test. For the TE-F group a significant difference was found ($\chi^2(2) = 6.465$, $p = .039$), however pairwise comparisons revealed no difference between the individual conditions. The overall significant difference can be accounted for by the TE-F group's higher scores on the A+I condition than the A+A or I+A conditions (Table 4.17), which may have been due to the fact that, for the A+I condition in SVO test items, both the word order and Animacy cues were aligned.

Comparing Animacy conditions (post-test)

At post-test, no differences were found between the learners' scores on any of the Animacy conditions (TE-FMC, Friedman's ANOVA, $\chi^2(2) = 0.667$, $p = .717$; TE-F, $\chi^2(2) = 0.500$, $p = 779$; Control, $\chi^2(2) = 1.400$, $p = .497$).

ii) OVS condition

Descriptive statistics for the OVS test items demonstrated that the three groups were performing at baseline at pre-test (*Mdn* = 0) across all Animacy conditions:

Table 4.18: Descriptive statistics for the Animacy conditions on the OVS test condition (Act-Out Comprehension)

Group	N	A+A (*k* = 3)					
		Pre-test			Post-test		
		M	**SD**	***Mdn***	***M***	**SD**	***Mdn***
TE-FMC	45	0.00	0.00	0	1.33	1.37	1
TE-F	41	0.12	0.33	0	1.10	1.18	1
Control	46	0.13	0.34	0	0.07	0.25	0
Total	132	0.08	0.28	0	0.82	1.18	0
Group	**N**	**A+I (*k* = 3)**					
TE-FMC	45	0.29	0.51	0	1.24	1.26	1
TE-F	41	0.29	0.64	0	1.41	1.32	2
Control	46	0.87	0.35	0	0.13	0.40	0
Total	132	0.22	0.51	0	0.91	1.21	0
Group	**N**	**I+A (*k*= 3)**					
TE-FMC	45	0.20	0.46	0	1.36	1.25	1
TE-F	41	0.23	0.52	0	1.27	1.26	1
Control	46	0.87	0.28	0	0.20	0.54	0
Total	132	0.17	0.43	0	0.92	1.17	0

Comparing Animacy conditions (pre-test)

There were no differences between the respective Animacy conditions for either the TE-F (Friedman's ANOVA, $\chi^2(2) = 1.480$, $p = .477$) or Control groups ($\chi^2(2) = 0.839$, $p = .657$). For the TE-FMC group an overall significant difference was found ($\chi^2(2) = 13.176$, $p = .001$); however pairwise comparisons revealed no differences between the individual conditions. The descriptive statistics (Table 4.18) indicated that the TE-FMC learners' performance was lower for the A+A condition than the A+I and I+A conditions.

Comparing Animacy conditions (post-test)

At post-test no differences were found between the Animacy conditions for either the TE-FMC (Friedman's ANOVA, $\chi^2(2) = 3.391$, $p = .183$) or Control groups ($\chi^2(2) = 3.765$, $p = .152$), whereas an overall significant difference was yielded for the TE-F group ($\chi^2(2) = 6.206$, $p = .045$). An examination of the descriptive statistics (Table 4.18) revealed that the TE-F group performance was highest on the A+I condition and lowest on the A+A condition, however this difference was not significant.

4.5.2 Gains made on the Animacy conditions (A+A, A+I, I+A)

Table 4.19 details each group's gains score[64] for the respective Animacy conditions sub-divided by test condition (SVO, OVS).

Table 4.19: Gains scores for the Animacy conditions by test condition (Act-Out Comprehension)

Group	N	A+A					
		SVO			OVS		
		M	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	45	-0.02	0.40	0	1.33	1.37	1
TE-F	41	-0.02	0.42	0	0.97	1.20	1
Control	46	0.11	0.57	0	-0.07	0.39	0
Total	132	0.03	0.47	0	0.73	1.22	0
Group	**N**	**A+I**					
TE-FMC	45	-0.07	0.25	0	0.96	1.54	1
TE-F	41	-0.02	0.42	0	1.12	1.58	2
Control	46	0.20	0.62	0	0.04	0.21	0
Total	132	0.04	0.47	0	0.69	1.34	0
Group	**N**	**I+A**					
TE-FMC	45	0.02	0.34	0	1.16	1.46	1
TE-F	41	0.13	0.60	0	1.04	1.39	1
Control	46	0.13	0.40	0	0.11	0.64	0
Total	132	0.09	0.45	0	0.75	1.29	0

i) SVO condition

It was hypothesised that the largest gains would be made on the I+A condition (hypothesis B). However analysis revealed that there were no differences in the gains made on the three Animacy conditions for any of the groups (TE-FMC, Friedman's ANOVA, $\chi^2(2) = 2.167$, $p = .338$; TE-F, $\chi^2(2) = 1.853$, $p = .396$; Control, $\chi^2(2) = 0.603$, $p = .740$). This finding was likely due to the fact that all of the learners were performing at ceiling level on the SVO test items from pre-test (section 4.4.1.2).

ii) OVS condition

Analysis found no differences for either the TE-F (Friedman's ANOVA, $\chi^2(2) = 2.606$, $p = .272$) or Control group ($\chi^2(2) = 2.440$, $p = .295$), in the gains made on the three Animacy conditions for OVS test items. For the TE-FMC group, however, an overall significant difference was found ($\chi^2(2) = 8.505$, $p = .014$). Pairwise comparisons revealed no differences between any of the

[64] Post-test score minus pre-test score

individual Animacy conditions; however the descriptive statistics (Table 4.19) indicated that the TE-FMC learners made the largest gains on the A+A condition, followed by the I+A and then the A+I condition.

In addition an examination of the descriptive statistics revealed that the TE-FMC and TE-F groups made substantially larger gains across all of the Animacy conditions on OVS test items when compared to the SVO test items, reflecting the learners' significant improvement on the OVS condition at post-test. In contrast the Control group gains across the Animacy conditions were minimal (for SVO and OVS), reflecting the lack of change in their performance between pre- and post-test.

4.5.3 Summary of findings for the Animacy conditions

Overall the findings presented above suggested that learners were not relying on Animacy cues when interpreting target language sentences. Firstly hypothesis A was not borne out; minimal differences were observed between the Animacy conditions (by total scores and test condition) and the learners' scores were not higher on the I+A condition for OVS test items at pre-test. Secondly, contrary to hypothesis B, there were no differences in the gains made by the three groups on any of the Animacy conditions between pre- and post-test for either the SVO or OVS test items.

It is also important to note that the descriptive statistics reflected the learners' predominant reliance on word order (particularly at pre-test) when interpreting the aural sentence stimuli, as would be expected for L1 speakers of English. Furthermore, the large gains made by the TE-FMC and TE-F groups (but not the Control group) across all of the Animacy conditions on the OVS test items were reflective of the improvement in the learners' comprehension of the target feature (section 4.4.1.2).

4.6 Performance on the metalinguistic task

4.6.1 *Sentence Reconstruction* task: Quantitative analysis

The quantitative analysis of the Sentence Reconstruction task (section 3.6.5) will be presented in two parts. The learners' scores for ordering the words in each sentence (Order) will be presented, followed by the scores relating to the explanations given (Explanation).

4.6.1.1 Analysis of scores for Order

Over time (pre- to post-test)

All three groups made significant improvement on the Order sub-task between pre- and post-test (TE-FMC, Wilcoxon signed rank, T = 571.00, p = .001, r = .41; TE-F, T = 413.50, p = .001, r = .43; Control, T = 396.00, p = .028, r = .23) (Table 4.20, Figure 4.12).

Table 4.20: Descriptive statistics for Sentence Reconstruction task – Order

Group	N	Pre-test (k = 6)			Post-test (k = 6)			Delayed post (k = 6)		
		M	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	45	4.53	1.2	4	5.62	0.83	6	5.33	1.13	6
TE-F	41	4.12	1.42	4	5.54	1.23	6	5.34	1.13	6
Control	46	3.89	1.42	4	4.48	1.43	4	-	-	-
Total	*	4.82	1.36	4	5.20	1.29	6	5.34	1.12	6

* Pre- and post-test, *N* = 132; Delayed post-test, *N* = 86

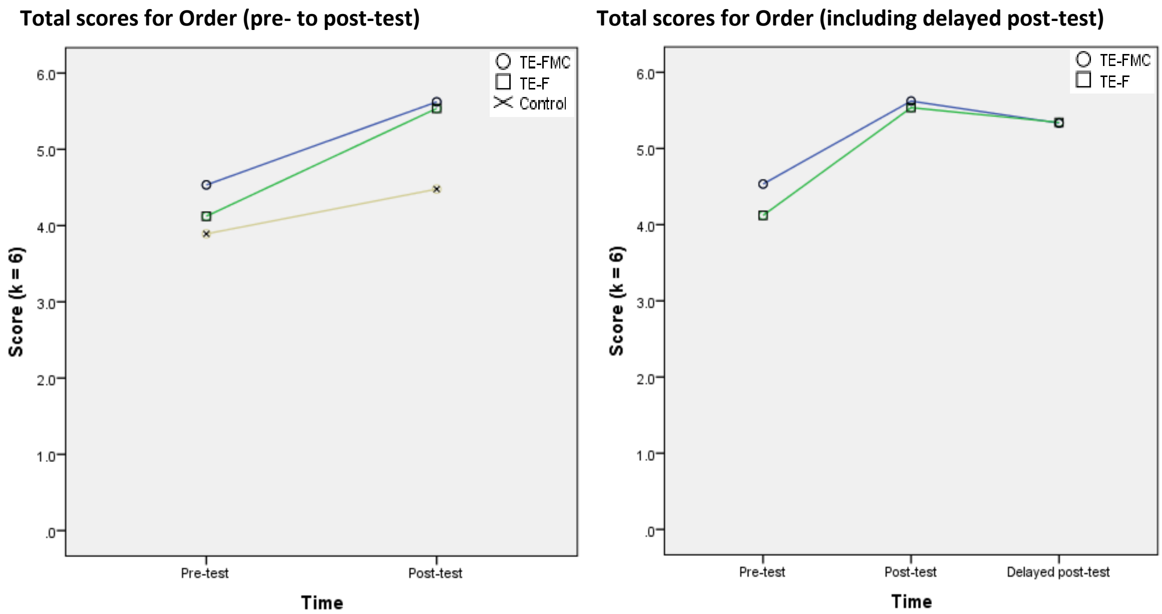


Figure 4.12: Total scores on the Sentence Reconstruction task (Order)

Over time (pre-, post- to delayed post-test)

Analysis revealed an overall significant change in both the TE-FMC (Friedman's ANOVA, χ^2 (2) = 20.588, *p* = .001) and TE-F (χ^2 (2) = 31.196, *p* = .001) group's scores over the three time points. For both groups, pairwise comparisons revealed an improvement between pre- and post-test (TE-FMC, *z* = -3.426, *p* = .001, *r* = -.36; TE-F, *z* = -4.196, *p* = .001, *r* = -.46) and pre- and delayed post-test (TE-FMC, *z* = -2.899, *p* = .004, *r* = -.31; TE-F, *z* = -3.258, *p* = .001, *r* = -.36) but no change between post- and delayed post-test (TE-FMC, *z* = 0.527, *p* = .598, *r* = .06; TE-F, *z* = 0.939, *p* = .348, *r* = .10).

Between groups (pre- and post-test)

The difference between the three groups approached significance at pre-test (Kruskall Wallis, *H*(2) = 5.75, *p* = .057), with the Control group performing at a slightly lower level than the TE-FMC or TE-F groups (Table 4.20, Figure 4.12). This difference was significant at post-test (*H*(2) = 27.34, *p* = .001), with both the TE-FMC (*p* = .001, *r* = .47) and TE-F groups (*p* = .001, *r* = .49) outperforming the Control group.

Between groups (delayed post-test)

No difference was found between the performance of the TE-FMC and TE-F groups at delayed post-test (Mann Whitney, $U = 923.500$, $z = 0.011$, $p = .991$, $r = .001$).

4.6.1.2 Analysis of scores for Explanation

The descriptive statistics for the learners' scores on the Explanation sub-task are provided in Table 4.21 and Figure 4.13:

Table 4.21: Descriptive statistics for Sentence Reconstruction task - Explanation

Group	N	Pre-test (k = 6)			Post-test (k = 6)			Delayed post (k = 6)		
		M	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	45	0.07	0.33	0	4.78	1.51	6	3.46	2.23	4
TE-F	41	0.00	0.00	0	4.32	2.02	5	2.90	1.84	3
Control	46	0.00	0.00	0	0.04	0.21	0	-	-	-
Total	*	0.02	0.19	0	2.99	2.59	4	3.19	2.06	4

* Pre- and post-test, $N = 132$; Delayed post-test, $N = 86$

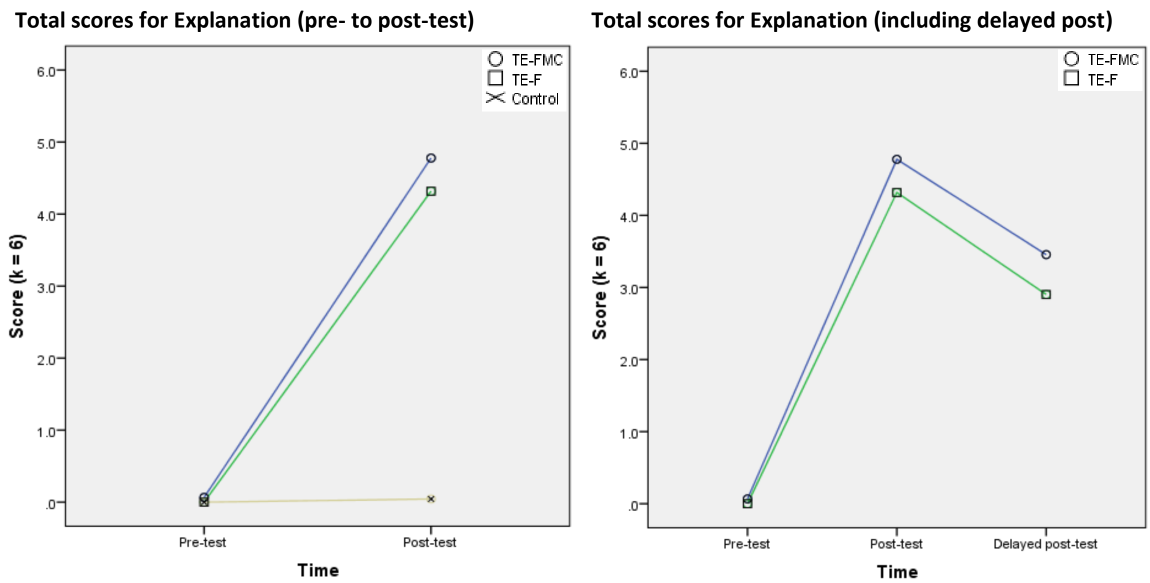


Figure 4.13: Total scores on the Sentence Reconstruction task (Explanation)

Over time (pre- to post-test)

A significant improvement was observed in both the TE-FMC (Wilcoxon signed rank, $T = 990.00$, $p = .001$, $r = .62$) and TE-F groups' scores ($T = 666.00$, $p = .001$, $r = .59$) between pre- and post-test. In contrast no change over time was found for the Control group ($T = 3.00$, $p = .157$, $r = .15$).

Over time (pre-, post- to delayed post-test)

Analysis revealed a significant change in the learners' scores across the three time points, for both the TE-FMC (Friedman's ANOVA, $\chi^2(2) = 65.790$, $p = .001$) and TE-F groups ($\chi^2(2) = 59.842$, $p = .001$). Pairwise comparisons revealed that this change was not only due to a significant increase in the learners' scores from pre- to post-test (TE-FMC, z = -7.537, p = .001, r = -.79; TE-F, z = -7.012, p = .001, r = -.77) and pre- to delayed post-test (TE-FMC, z -5.112, p = .001, r = -.54; TE-F, z = -4.583, p = .001, r = -.51), but also due to a significant decrease in their performance between post- and delayed post-test (TE-FMC, z = 2.429, p = .015, r = .26; TE-F, z = 2.429, p = .015, r = .27) (Table 4.21, Figure 4.13).

Between groups (pre- and post-test)

No difference was found between the three groups at pre-test (Kruskall Wallis, $H(2)$ = 3.90, p = .143). At post-test, a significant difference ($H(2)$ = 86.13, p = .001) was found between the Control group and both the TE-FMC (p = .001, r = .89) and TE-F groups (p = .001, r = .79).

Between groups (delayed post-test)

There was no difference between the TE-FMC and TE-F groups' performance on the Explanation sub-task at delayed post-test (U = 748.000, z = -1.527, p = .127, r = -.16).

4.6.1.3 Examination of effect sizes

The magnitude of instructional effect for the TE-FMC intervention was large for both the Order and Explanation sub-tasks (Table 4.22). Similarly the effect size calculated for the TE-F intervention in comparison to the Control group was medium to large for the Order sub-task. The large magnitude of change for the TE-FMC group on the Explanation sub-task reflected the learners' significant gains at post-test. Notably, the effect size is comparatively smaller (although still large) for the learners' pre- to delayed post-test performance, in line with the decrease in learners' scores observed between post- and delayed post-test.

Table 4.22: Magnitude of instructional effect on the Sentence Reconstruction task

Contrast	Group(s)	Order	Explanation
		d	***d***
Between groups	**TE-FMC vs. Control**	1.01	-[a]
	TE-F vs. Control	0.79	-[a]
Pre to post	**TE-FMC**	1.07	5.12
	TE-F	1.07	-[a]
	Control	0.41	-[a]
Pre to delayed post	**TE-FMC**	0.69	2.65
	TE-F	1.05	-[a]

[a]Effect sizes could not be calculated as SD = 0 for TE-F group at pre-test and Control group at pre- and post-test

4.6.1.4 Summary of findings for the quantitative analysis

At pre-test, all three groups were able to successfully complete the Order part of the Sentence Reconstruction task to a certain extent and the learners' accuracy significantly increased at post-test for all three groups. The increase in the Control group scores could indicate that the observed improvement may have been due to an increase in test familiarity at post-test.

However it is important to note that whilst all three groups improved, the TE-FMC and TE-F groups made substantially more progress than the Control group (Figure 4.12). Therefore it is likely that some, if not a majority, of the TE-FMC and TE-F group improvement at post-test was a result of the knowledge gained through the interventions.

The analysis of the learners' Explanation scores indicated that following their respective interventions, the TE-FMC and TE-F learners were able to articulate their knowledge of the target feature and its use within target language sentences. In contrast no improvement was observed in the Control group. Notably, the TE-FMC and TE-F learners' scores for the Explanation sub- task significantly decreased at delayed post-test. This finding was markedly different from the results of the written and oral tasks, since for all other tasks the learners had maintained their gains at delayed post-test (see section 6.3.2 for discussion).

4.6.2 *Sentence Reconstruction* task: Qualitative analysis

The learners' explanations were examined in order to better understand the extent of their metalinguistic knowledge in relation to the target grammatical feature. The following sections will present examples of the explanations given by the learners at the three time points and explore the nature of the change in metalinguistic knowledge exhibited by the TE-FMC and TE-F groups at post- and delayed post-test.

4.6.2.1 Metalinguistic knowledge at pre-test

i) Test item: two masculine nouns

At pre-test, none of the explanations given by the learners related to the function of ***der*** (subject) and ***den*** (object) in assigning thematic roles within the sentence. This was expected given that the learners had received no exposure or instruction relating to the target feature prior to the research study. Despite this, a majority of the participants were able to identify ***den*** as a type of article (either another word for ***the*** or as the indefinite article ***a***), for example:

> **P[65]:** The man wrote the letter. I'm not sure if that one (*den*) could go there (*before Brief*).
> **R[66]:** Ok, so why could that one go that way round, den Brief? What could den mean?
> **P:** a, a letter
>
> (Participant 22, TE-FMC, School 2)

> **R:** And why did you put der with Mann and den with Brief?
> **P:** Because Brief means letter and in English we would say the letter or a letter so den would go next to it. And der Mann, because in German der is a masculine word.
>
> (Participant 34, TE-F, School 2)

---

[65] **P** participant
[66] **R** researcher

Additionally a common explanation which arose was related to animacy; namely that ***den*** was for inanimate items (i.e. things), whereas ***der***, ***die***, and ***das*** were for animate nouns (i.e. humans or animals), for example:

> **R:** Ok, so we've put der Mann and den Brief, and you were thinking about swapping them (*der* and *den*) around. What made you put them this way round?
> **P:** I think that a person was der and a thing was den.
>
> (Participant 12, TE-FMC, School 2)

> **R:** Ok and how did you know that *den* would go with Ball?
> **P:** Because you use that for a thing.
>
> (Participant 112, Control, School 3)

Furthermore, many of the learners reported relying on intuition or guesswork when deciding on the position of the articles, for example:

> **R:** Ok, so why have you decided that den and Vogel would go together (.) and der and Hund?
> **P:** Well (.) I didn't really (.) I just tried to remember which was which and then took a guess.
>
> (Participant 75, TE-F, School 1)

> **R:** Ok… and why did you choose der with Mann? (..) Do you know what der means?
> **P:** the
> **R:** the, alright. And why did you choose to put der with Mann? Any reason?
> **P:** It doesn't sound right with den.
>
> (Participant 5, TE-FMC, School 1)

ii) Test item: one masculine plus one feminine or neuter noun

Many of the explanations given by the learners also related to gender, for example:

> **P:** der is the masculine (..) word for the and I think Hund is masculine, verfolgt means chase and the dog is chasing the cat, and then die is the (.) feminine thingy or um (.) feminine word for the, and Katze I think is a feminine word
>
> (Participant 17, TE-FMC, School 2)

> **R:** How did you know *der* would go with Vater?
> **P:** Because it's masculine
> **R:** Masculine ok. And what about *das* and Baby?
> **P:** Neutral
>
> (Participant 111, Control, School 3)

The participants across the three groups were comfortable with the abstract concept of gender and as demonstrated above, a majority were able to employ metalinguistic terms in their explanations. Some learners, however, relied on more colloquial terms, for example:

> **P:** Because um (.) der wouldn't go with Frau because (.) der is for male and (.) die is for female.
>
> (Participant 45, TE-FMC, School 1)

R: Ok so why do *der* and Mann go together?
P: […] Because, is *der* for a boy and *die* for a girl

(Participant 127, Control, School 2)

An additional strategy discussed by some learners was making use of morphological cues to help them identify which article belonged to each noun, for example:

P: Because I know that Hund is masculine and Katze is feminine. And the way I know it is Katze has an e and all feminine words have an e at the end.

(Participant 2, TE-FMC, School 1)

P: Katze has got an e on the end so it must be a feminine word.

(Participant 28, TE-F, School 1)

The above extracts are representative of the explanations given by the learners from all three groups at pre-test. Additionally explanations relating to gender, animacy, and guesswork were characteristic of the Control group explanations at post-test.

4.6.2.2 Metalinguistic knowledge at post-test

i) Test item: two masculine nouns

Following their respective teaching interventions, a majority of participants in both the TE-FMC and TE-F groups were able to correctly articulate the function of ***der*** and ***den*** in assigning subject and object roles within a sentence, for example:

P: That's (*der*) the subject so that would be the thing doing the action and that (*den*) would be the thing receiving the action.
R: Ok so der is the subject and den
P: That's the object.

(Participant 16, TE-FMC, School 2)

P: Because I knew that der is for the subject of the sentence, the thing that does the action. And den is for the object, the thing being done to. And the dog is being chased by the bird. So der Vogel verfolgt den Hund, and Hund is dog.

(Participant 33, TE-F, School 2)

As demonstrated in the extracts above, a majority of the learners were able to employ the metalinguistic terms subject and object in their explanations. In contrast, whilst some of the learners avoided using these terms, they were still able to explain the role of the articles ***der*** and ***den*** in their own words, for example:

P: Um because I know the Vogel was a bird and it was chasing the dog so I put *der* there in front of Vogel and it was chasing um (.) and then the dog is being chased so it's den Hund.

(Participant 14, TE-FMC, School 2)

> **P:** Because I knew that <u>der</u> means the person (.) the thing that's doing the action and <u>den</u> means the person that's receiving the action.
>
> (Participant 70, TE-F, School 1)

ii) Test item: one masculine plus one feminine or neuter noun

One article for feminine and neuter nouns

Many of the learners were able to articulate that, in contrast to masculine nouns, feminine and neuter nouns only use one article (***die*** and ***das*** respectively) for both the subject and object of the sentence, for example:

> **R:** And is there anything else you can tell me about <u>das</u> maybe?
> **P:** Well there are not two different words for the subject and the object, <u>das</u> is the word for both
>
> (Participant 24, TE-F, School 1)

> **R:** Ok… do you know anything else about <u>die</u>?
> **P:** It doesn't have an alternate word that means the same thing for object, like <u>der</u> has <u>den</u>, it's just <u>die</u> and <u>die</u>.
>
> (Participant 2, TE-FMC, School 1)

This difference between masculine versus feminine and neuter nouns had been briefly explained to the TE-FMC and TE-F learners' as part of the explicit information provided during weeks 4 and 5 of the intervention. Many learners were able to draw on this knowledge in their explanations, as demonstrated above.

Crucially the learners were able to extrapolate their knowledge of masculine articles and by a process of elimination were able to work out the function of the 'non-case-marked' feminine or neuter article, for example:

> **P:** I mean <u>die</u> is a feminine noun and <u>den</u> is (.) used for object, masculine. And <u>die</u> can be used for subject and object. But because <u>den</u> is used for the object, then <u>die</u> will be used for the subject of the sentence.
>
> (Participant 50, TE-FMC, School 1)

> **P:** Well the kid is hugging the teddy bear and <u>den</u> is um (.) the masculine word that's used as the object, so I thought <u>das</u> must be the subject since <u>den</u> is the object.
>
> (Participant 25, TE-F, School 1)

iii) Word order

At post-test a number of learners also took the opportunity to demonstrate their understanding of word order in German, explaining that due to the presence of ***der*** and / or ***den*** it was possible to reverse the order of a German sentence without changing the meaning, for example:

P: Because der is the (.) subject, der is to describe what the subject is. And das is to describe what the object is. (.) Or you could do (.) um it that way round. (*Participant swaps der Vater and das Baby*)
R: ok, das Baby küsst der Vater. Why can you have it that way round?
P: You can have it that way round because (.) you'll still know which way round it goes (.) because der is the subject (.) and das is the object.

(Participant 59, TE-FMC, School 2)

P: At first I had it the other way round but then I swapped it.
R: Ok so why did you swap it round this way?
P: Because even if it's this order it still means the same thing.
R: Ok so in German the word order can change.
P: Because that (*den*) is still the object

(Participant 2, TE-FMC, School 1)

These learners demonstrated that they were willing to circumvent the word order rule from their L1 and could correctly interpret OVS sentences using the masculine articles.

Nevertheless, there was also evidence that some of the learners continued to rely on their L1 word order rule of subject first object second when constructing and interpreting sentences, for example:

P: Yeah so (.) um the ball is hitting the football player on the head. And Ball is a masculine noun but so is Fuβballspieler, but you put den at the end because Fuβballspieler is the (.) object of the sentence. And der is the subject.

(Participant 48, TE-FMC, School 1)

P: Because der is the subject so it goes first. And it is the bird because it's doing the chasing. And den is the object so it goes last and then it's the Hund.

(Participant 39, TE-F, School 2)

Despite correctly articulating the role of ***der*** and ***den*** in assigning subject and object, the extracts above demonstrated that some learners continued to associate the subject with the 'first thing' and the object with the 'second thing' in the sentence.

Finally it is important to note that at post-test there were a small number of learners who continued to rely on explanations relating to gender, animacy, and guesswork (as at pre-test) and were unable to articulate the grammatical role-assigning function of ***der*** and ***den***.

4.6.2.3 Metalinguistic knowledge at delayed post-test

i) Improvement between pre- and delayed post-test

At delayed post-test, as at post-test, there was evidence of the learners' having developed metalinguistic knowledge of the target feature, which was not observed at pre-test. Many learners were able to identify the function of ***der*** and ***den*** in assigning subject and object roles within the sentence, as well as apply this knowledge to sentences containing a feminine or neuter article. Additionally some learners explained the role of the target feature when interpreting German sentences in reversed word order as at post-test.

ii) Inconsistencies between post- and delayed post-test

Despite the overall improvement observed from pre-test, the explanations given by the learners at delayed post-test suggested that their metalinguistic knowledge was less reliable than it had been at post-test, for example:

Post-test

P: […] So I know <u>den</u> is for the object so I know this (*<u>das</u>*) is going to be the subject. And the kid is the subject because he's doing (.) it's cuddling the teddy.

Delayed post-test

P: Well the father is kissing the baby. These (*<u>der</u>* and *<u>das</u>*) (.) <u>das</u> can either go at the start or at the end, because if it's <u>die</u> or <u>der</u> (.) I think (.) they go at the start. But if it's uh (.) I can't remember the other one (.) <u>den</u> or something, then that one (*<u>das</u>*) goes at the start.

(Participant 15, TE-FMC, School 2)

Post-test

P: Because (.) the clown is doing the scaring and the woman is scared on the picture. And <u>der</u> is the subject and <u>die</u> is the object.

R: […] And anything else you can tell me about <u>die</u>?

P: <u>die</u> is the feminine word. But in sentences <u>die</u> is also used for the subject and the object.

Delayed post-test

R: So it's masculine exactly, so Computer goes with <u>den</u> and <u>die</u> is feminine so it goes with Frau. And you said we have two masculine words, what are the two words we've seen?

P: <u>der</u> and <u>den</u>.

R: <u>der</u> and <u>den</u>, and what's the difference between them? Why do we have <u>den</u> in this one?

P: <u>den</u> means, just like a normal the. But <u>der</u> it could be (.) well he, or something like that.

(Participant 34, TE-F, School 2)

In both extracts, at post-test, the learners were able to articulate the function of the masculine article in the sentence and apply this to work out the role of the feminine or neuter article. However, at delayed post-test, neither learner was able to provide an explanation for the role of the articles in the sentence.

iii) Inconsistencies between test items

Additionally, at delayed post-test, many learners were able to give correct explanations for one test item, however were unable to do so on the next test item (or vice versa). This pattern was observed for many of the learners, for example:

Test item 1

P: Um (…) I knew that (..) it was the father kissing the baby. And (..)

R: You could see from the picture. And what about the words, I guess you knew Baby, and Vater is

P: father.

R: What about <u>der</u> and <u>das</u>?

P: Um I knew that Vater was masculine.

R: Ok so you put <u>der</u> there (*before Vater*)

P: And I knew that Baby was neuter.

Test item 3

P: Well um (.) <u>der</u> is the subject and <u>den</u> is the object, and it doesn't matter which way round you put them in the sentence. But the ball is (.) hitting the (.) football player, so (.) um so the ball is the subject.

(Participant 9, TE-FMC, School 1)

When the learners gave the correct explanation, as in test item 3 above, they were often able to utilise the correct metalinguistic terms. However when the learners failed to do so, their explanations tended to be centred on gender, as in test item 1 above.

Notably, at delayed post-test, correct explanations were more consistently provided for the test items containing two masculine nouns. For the test items containing one masculine plus one feminine or neuter noun the learners' metalinguistic knowledge appeared to be less reliable, for example:

Test item 2

P: Well <u>die</u> goes with woman (*Frau*) and she's doing the action so it goes with her. And Frau means lady. And schlägt means hit, and <u>den</u> means the. And Computer.

R: the computer. You said that <u>die</u> and Frau go together, because it's a woman. How about <u>den</u> and Computer? Why do those two fit together in this sentence?

P: Because um (…)

R: Any ideas? (.) or just thought it sounded right?

P: Yes.

Test item 3

P: <u>der</u> goes with the thing that's acting out the action. And the ball hit the Fußballspieler. And <u>den</u> goes with Fußballspieler because he's receiving the action.

(Participant 14, TE-FMC, School 2)

This learner was unable to articulate the relationship between ***den*** and *Computer* in test item 2; however the same learner correctly explained that ***den*** identified the noun that was receiving the action in test item 3.

Additionally, many of the learners failed to extrapolate their knowledge of the masculine articles ***der*** and ***den*** to the non-case-marked feminine or neuter articles at delayed post-test, for example:

P: Um (.) well it's the father doing the action so it has to be der.
R: Ok the father is doing the action so it has to be der.
P: And then that leaves das so I put that with the baby (.) and the father is kissing the baby.

(Participant 32, TE-F, School 1)

It was these inconsistencies in the explanations given by the participants at delayed post-test which accounted for the significant decrease in scores found via the statistical analysis (section 4.6.1.2).

4.6.2.4 Summary of findings for the qualitative analysis

The qualitative data provided evidence of the change in the learners' metalinguistic knowledge over the three time points. An examination of the explanations given at pre-test confirmed that the learners, across all three groups, had no metalinguistic knowledge of the target feature (in terms of assigning subject and object roles), as was the case for the Control group at post-test. At post-test, however, the learners from both the TE-FMC and TE-F groups demonstrated that they were able to correctly articulate their knowledge of the target feature, either in their own words or by making use of appropriate metalinguistic terminology. Additionally, the learners were able to apply their knowledge of masculine articles in order to aid interpretation of sentences containing a non-case-marked feminine or neuter article, as well as sentences in reversed word order. In contrast, at delayed post-test, inconsistencies were found in the explanations given by the learners for different test items, as well as inconsistencies between the explanations given at post- and delayed post-test.

4.7 Performance on the six outcome measures

4.7.1 Overall magnitude of instructional effect

Both the TE-FMC and TE-F interventions yielded large overall (mean) effect sizes in comparison to the Control group:

Table 4.23: Overall magnitude of instructional effect

Contrast	Group(s)	k^{ab}	Mean *d*	SD *d*	95% CI (lower)	95% CI (higher)
Between groups	TE-FMC vs. Control	6	1.20	0.48	0.80	1.60
	TE-F vs. Control	6	0.94	0.75	0.14	1.55
Pre to post	TE-FMC	6	1.63	0.44	1.27	1.99
	TE-F	6	1.60	0.59	1.12	2.08
	Control	6	0.20	0.25	0.05	0.40
Pre to delayed post	TE-FMC	6	1.63	0.53	1.20	2.06
	TE-F	6	1.54	0.44	1.18	1.99

[a] Number of dependent variables contributing effect sizes
[b] Excluding Sentence Reconstruction (Explanation) since effect size could only be calculated for TE-FMC group

Furthermore, the TE-FMC and TE-F interventions resulted in large, equivalent gains between both pre- and post-test and pre- and delayed post-test. These findings reflected the significant improvement in both groups' scores at post- and delayed post-test across all tasks. In contrast, the mean effect size calculated for the Control group between pre- and post-test was small, reflecting the lack of change in the Control group's performance between pre- and post-test on any of the tasks.

4.7.2 Summary of main findings

This chapter has examined the learners' performance on each task both over time (pre-, post-, delayed post-test) and between groups (TE-FMC, TE-F, Control).

The preliminary analysis of the learners' pre-test performance by age and school established that the participants were performing at an equivalent level at baseline. In terms of the learners' vocabulary knowledge, the TE-F and Control groups were found to be performing at a slightly lower level than the TE-FMC group at pre-test. However at post-test the TE-F and Control groups' scores had increased and no difference was found between the three groups. Further the difference in the TE-FMC versus TE-F and Control groups' vocabulary knowledge at pre-test did not appear to impact their performance on the outcome measures, since no differences were found between the three groups at pre-test on any of the tasks, nor between the TE-FMC and TE-F groups at post- or delayed post-test.

The inferential statistical analysis of the three groups' performance on the outcome measures revealed a clear pattern in the data. Across both the written and oral tasks, substantial improvement was made by both the TE-FMC and TE-F groups at post-test, which was sustained at delayed post-test, nine weeks after the intervention. Notably the learners received no extra exposure to or instruction on the target feature between post- and delayed post-test, therefore this continued high level of performance was likely a lasting effect of the respective instructional treatments, which the two groups had received.

In contrast, no change was observed in the Control group's scores. This finding of 'no change' for the Control group suggested that the improvement made by the TE-FMC and TE-F groups was not simply a test effect, i.e. the two groups had not improved at post-test simply because they had completed the tasks once before and were familiar with the task format from the pre-test.

Furthermore, the effect sizes calculated demonstrated that not only did the TE-FMC and TE-F groups improve at post-test, they had made *substantial* gains on all of the tasks. This was further demonstrated in the analysis conducted on the data for each test condition. All three groups were found to be consistently performing at ceiling level on the SVO / subject focussed conditions (i.e. distractor items). Consequently the gains made by the TE-FMC and TE-

F groups at post-test (and maintained at delayed post-test) can be almost exclusively attributed to an improvement in their comprehension of the target feature (***den***) within sentences presented in reversed word order (OVS) and / or their accuracy in producing the target feature within both SVO and OVS sentences.

Finally, the Sentence Reconstruction task did not follow the same pattern of results as that of the written and oral outcome measures. Crucially, whilst the TE-FMC and TE-F groups had made gains on the Explanation sub-task at post-test, both groups' scores had significantly decreased at delayed post-test. This finding indicated that the learners' ability to articulate their metalinguistic knowledge of the target feature had substantially lessened by delayed post-test.

Chapter 5: Results 2

The first part of this chapter will present the results of the Principal Component Analysis (PCA), which explored the relationships between the six outcome measures at post- and delayed post-test (section 5.1). Section 5.2 examines individual participant performance across the six outcome measures. Finally section 5.3 explores the potential confounding effect of grammatical sensitivity on the learners' performance. An initial interpretation of the results of the analysis will be included in this chapter (for full discussion see chapter 6).

5.1 Relationship between the outcome measures: Principal Component Analysis

The learners' performance across the six outcome measures was compared in order to explore the type of knowledge elicited by each of the tasks following the TE-FMC and TE-F interventions. It was predicted that the written activities (Sentence Matching, Gap-fill) would elicit more explicit knowledge and the oral tasks (Sentence Repetition, Act-Out Comprehension, Act-Out Production) a more implicit form of knowledge (see section 3.6). Additionally the Explanation element of the Sentence Reconstruction task constituted a test of the learners' metalinguistic knowledge. As in previous studies (R. Ellis, 2005; Han & Ellis, 1998; Marsden & Chen, 2011), PCA was utilised in order to examine the pattern underlying participant performance across the six tasks and investigate the predictions made regarding the type of knowledge likely to be elicited.

5.1.1 Preparing the data for PCA

i) Datasets to be included

Given that the aim of the PCA was to examine the types of knowledge developed following the teaching interventions (TE-FMC and TE-F), the Control group data was not included (*n* = 52). Additionally those TE-FMC and TE-F learners, who had not completed all of the tasks at post- and delayed post-test, were excluded from the analysis (*n* = 4). Given that, universally, no difference was found between the two groups on any of the tasks at any of the time points, the PCA was run with the learners as one group (All learners). However, analysis was also run on the TE-FMC and TE-F group data separately, in order to ascertain whether there were any differential effects for the two intervention types. Additionally only the data for the Explanation element of the Sentence Reconstruction task was included in the PCA, given that it was this part of the activity which was designed to test metalinguistic knowledge.

ii) *Suitability of the datasets for PCA*

The suitability of the respective datasets for analysis via PCA was determined by an examination of the following categories (see section 3.7.3.1):

a) Number of cases
b) Kaiser-Meyer-Olkin (KMO) statistic
c) Bartlett's test of sphericity
d) *R* determinant of the *R*-matrix

Table 5.1 details the information for each of these categories:

Table 5.1: Suitability of data for PCA

Dataset	*N*	Number of Cases	Time	KMO	*R*	Bartlett's test χ^2	Bartlett's test *p*
All learners	82	492	Post	0.706	.116	168.233	.001
			Delayed	0.825	.061	219.197	.001
TE-FMC	44	264	Post	0.619	.137	79.869	.001
			Delayed	0.687	.048	121.664	.001
TE-F	38	228	Post	0.738	.052	100.945	.001
			Delayed	0.791	.031	118.443	.001

All three datasets satisfied these criteria. Firstly the number of cases in the 'All learners' dataset was sufficiently large (> 300). Although the number of cases within the TE-FMC and TE-F datasets was below this threshold, the number of participants was above the recommended minimum (10 to 15). Secondly the KMO statistics yielded were either mediocre (0.5 < KMO < 0.7), good (0.7 < KMO < 0.8) or very good (0.8 < KMO). Thirdly Bartlett's test of Sphericity revealed that the correlations between variables were sufficiently large within each dataset. Finally the *R* determinants (> .00001) indicated that the datasets did not suffer from multicollinearity[67].

5.1.2 PCA of full dataset (All learners)

i) *Analysis of post-test data*

The maximum number of points available varied between tasks; therefore the learners' raw scores for each task were converted into percentages in order to directly compare the activities (Table 5.2. At post-test the learners were performing at a higher level on the written and metalinguistic tasks than on the oral tasks, with the highest score being for the Sentence Matching task (written, comprehension) and the lowest for the Act-Out Comprehension task (oral, comprehension).

[67] Very highly correlated variables

Table 5.2: Percentage score for each task (All learners)

Task	*N*	Post-test		Delayed post-test	
		Mean %	*SD*	*Mean %*	*SD*
S. Matching	82	81.35	19.53	78.51	19.82
Gap-fill	82	74.80	22.72	71.39	21.25
S. Repetition	82	70.83	16.88	72.87	18.23
Act-Out Comp	82	69.51	19.43	70.19	19.02
Act-Out Prod	82	70.27	23.48	73.88	24.15
S. Recon (E)	82	76.83	28.42	54.78	33.46

Table 5.3 presents the correlation matrix (calculated using Pearson's correlation) for the learners' scores across the six tasks. A significant ($p < .05$ level or higher) level of intercorrelation was observed between all pairs of tests, with the exception of the Sentence Reconstruction (Explanation) task, which correlated with the Sentence Matching task ($r = .25$, $p = .011$) only.

Table 5.3: Correlation matrix for six outcome measures (All learners)

Time	Task	S. Match	Gap-Fill	S. Rep	AO Comp	AO Prod	S. Recon (E)
Post-test	**S. Match**	-	.68**	0.28**	.32**	.39**	.25*
	Gap-Fill		-	.33**	.40**	.49**	0.13
	S. Rep			-	.41**	.48**	-0.01
	AO Comp				-	.77**	0.09
	AO Prod					-	0.11
	S. Recon (E)						-
Delayed post-test	**S. Match**	-	.51**	.42**	.56**	.59**	.45**
	Gap-Fill		-	.27**	.41**	.51**	.25*
	S. Rep			-	.60**	.64**	.40**
	AO Comp				-	.81**	.44**
	AO Prod					-	.42**
	S. Recon (E)						-

*significant at the .05 level
**significant at the .01 level

The PCA extracted two underlying components from the data set; the eigenvalues[68] for the two components were greater than 1 and therefore satisfied Kaiser's criterion, which states that extracted components with an eigenvalue greater than 1 should be included within a factor solution, as they are likely to represent a substantial amount of variation (Field, 2009).

Table 5.4 presents the factor loadings for each of the outcome measures onto the two extracted components. The three oral tasks were found to load heavily (> 0.7) onto component 1, whereas the written and metalinguistic tasks loaded onto component 2.

[68] Post-test: Component 1, eigenvalue of 2.870 (47.83% variance); Component 2, eigenvalue of 1.127 (18.79% variance).

Table 5.4: Factor loading for each task onto extracted component(s) (All learners)

Task	Post-test		Delayed post-test
	1	**2**	**1**
S. Matching		0.699	0.775
Gap-fill		0.527	0.630
S. Repetition	0.743		0.738
Act-Out Comp	0.842		0.860
Act-Out Prod	0.868		0.892
S. Recon (E)		0.800	0.631

ii) Analysis of delayed post-test data

At delayed post-test the learners' highest score was again for the Sentence Matching task (Table 5.2). The lowest score, however, was for the Sentence Reconstruction (Explanation) task, reflecting the significant decrease in the learners' scores on this task (section 4.6.1.2).

The correlation matrix for the delayed post-test data (Table 5.3) revealed significant correlations between all of the tasks, including the Sentence Reconstruction task.

In contrast to the results of the PCA for the post-test data, only one underlying component[69] was extracted from the delayed post-test data (Table 5.4). All of the variables loaded strongly onto this component.

5.1.3 PCA of group datasets (TE-FMC, TE-F)

i) Analysis of post-test data

Table 5.5 presents the overall percentage scores for the TE-FMC and TE-F groups on each measure. For both groups, the highest score was on the Sentence Matching task. Overall, performance was lower on the oral tasks than on the written and metalinguistic tasks.

Table 5.5: Percentage score for each task (by group)

Task	Post-test				Delayed post-test			
	TE-FMC[a]		TE-F[b]		TE-FMC		TE-F	
	Mean %	*SD*	*Mean %*	*SD*	*Mean %*	*SD*	*Mean %*	*SD*
S. Matching	80.9	20.5	81.9	18.6	77.7	19.4	79.5	20.5
Gap-fill	73.8	22.7	76.0	23.0	70.9	20.9	71.9	21.9
S. Repetition	73.1	18.4	68.2	14.7	76.0	19.2	69.3	16.6
Act-Out Comp	69.7	20.8	69.3	18.0	71.1	19.9	69.2	18.2
Act-Out Prod	71.3	23.6	69.1	23.6	75.3	23.6	72.3	25.0
S. Recon (E)	76.2	25.2	74.1	31.9	58.9	36.5	50.0	29.3

[a] TE-FMC, N = 44
[b] TE-F, N = 38

The correlation matrix, sub-divided by experimental group (TE-FMC, TE-F) can be seen in Table 5.6:

[69] Delayed post-test: Component 1, eigenvalue of 3.476 (57.93% variance)

Table 5.6: Correlation matrix for six outcome measures (by group)

Time	Task	S. Match		Gap-Fill		S. Rep		AO Comp		AO Prod		S. Recon_E	
		TE-FMC	TE-F	TE-FMC	TE-F	TE-FMC	TE-F	TE-FMC	TE-F	TE-FMC	TE-F	TE-FMC	TE-F
Post-test	**S. Match**	-	-	0.62**	.75**	0.24	.37*	0.23	.46**	0.23	.60**	0.13	.40**
	Gap-Fill			-	-	0.21	.55**	.29*	.55**	.32*	.67**	0.12	0.14
	S. Rep					-	-	0.37**	.49**	.49**	.47**	-0.10	0.70
	AO Comp							-	-	.80**	.74**	0.10	0.07
	AO Prod									-	-	-0.01	0.21
	S. Recon (E)											-	-
Delayed post-test	**S. Match**	-	-	.56**	.46**	.39**	.49**	.50**	.65**	.44**	.75**	.44**	.51**
	Gap-Fill			-	-	0.24	.34*	.30*	.54**	.47**	.57**	.31*	0.19
	S. Rep					-	-	.63**	.55**	.72**	.55**	.39**	.41**
	AO Comp							-	-	.78**	.85**	.54**	.29*
	AO Prod									-	-	.35**	.51**
	S. Recon (E)											-	-

*significant at the .05 level
**significant at the .01 level

For the TE-FMC group a significant level of intercorrelation ($p < .01$) was observed between the two written tasks and also between the three oral tasks, whereas the Sentence Reconstruction (Explanation) task was not found to significantly correlate with any of the activities. For the TE-F group significant correlations were observed between all of the written and oral tasks. In addition the Sentence Reconstruction task correlated significantly with the Sentence Matching task ($r = .395$, $p = .007$).

Two underlying components[70] were extracted from the datasets of both groups (Table 5.7); however, the pattern of factor loadings differed between the two groups.

Table 5.7: Factor loading for each task onto extracted component(s) (by group)

Task	**Post-test**				**Delayed post-test**	
	TE-FMC		**TE-F**		**TE-FMC**	**TE-F**
	1	**2**	**1**	**2**	**1**	**1**
S. Matching		0.807	0.633		0.725	0.844
Gap-fill		0.784	0.837		0.605	0.664
S. Repetition	0.721		0.756		0.768	0.711
Act-Out Comp	0.838		0.853		0.858	0.860
Act-Out Prod	0.915		0.847		0.859	0.930
S. Recon (E)		0.531		0.949	0.654	0.599

For the TE-FMC group, the three oral tasks loaded heavily onto Component 1, whereas the written and metalinguistic tasks loaded onto component 2. For the TE-F group, however, the written *and* oral tasks were found to load onto component 1, and the Sentence Reconstruction task loaded separately onto component 2.

ii) Analysis of delayed post-test data

Table 5.5 (above) details the TE-FMC and TE-F groups' percentage scores on each task at delayed post-test. For both groups, the lowest score was observed for the Sentence Reconstruction (Explanation) task, which was in line with the significant decrease observed in the learner's performance on this task at delayed post-test (section 4.6.1.2).

A significant intercorrelation was observed between all of the tasks for the TE-FMC group (Table 5.6), except between the Gap-Fill and Sentence Repetition tasks ($r = .237$, $p = .061$). Likewise for the TE-F group all of the tasks were found to significantly correlate, except the Gap-Fill and Sentence Reconstruction tasks ($r = .193$, $p = .123$).

For both the TE-FMC and TE-F groups only one component[71] was extracted from the delayed post-test dataset (Table 5.7); all of the tasks loaded heavily onto this component.

[70] **TE-FMC**: Component 1, eigenvalue of 2.549 (42.49% variance); Component 2, eigenvalue of 1.282 (21.37% variance). **TE-F**: Component 1, eigenvalue of 3.343 (55.71% variance); Component 2, eigenvalue of 1.072 (17.87% variance)

5.1.4 Summary of PCA findings

At post-test, two components were extracted from the full dataset (All learners), with the oral tasks loading onto component 1 and the written and metalinguistic tasks onto component 2. This finding suggested that the written and metalinguistic tasks may have been tapping into a different type of knowledge (i.e. more explicit) to that of the oral tasks (i.e. more implicit). The same pattern was observed in the analysis of the TE-FMC post-test dataset. For the TE-F group however the only task to load onto component 2 was the Sentence Reconstruction task. Given that the Sentence Reconstruction task was a test of metalinguistic (explicit) knowledge, this finding indicated that the TE-F group may have been relying less on their metalinguistic knowledge when completing both the written and oral tasks at post-test.

In contrast the analyses of the two groups' delayed post-test datasets were consistent with the analysis of the learners as one group; all of the tasks were found to load onto one component. The loading of the six tasks onto one component suggested that at delayed post-test all of the tasks were tapping into one type of knowledge. Coupled with the decrease in metalinguistic knowledge observed on the Sentence Reconstruction task at delayed post-test, this finding suggested that the learners may have been relying on a more implicit type of knowledge when completing the activities at this time point (see section 6.3.3 for discussion).

5.2 Individual participant performance on the outcome measures

The analysis presented in Chapter 4 revealed an overall significant improvement across all of the outcome measures for both the TE-FMC and TE-F groups. However at both post- and delayed post-test a large standard deviation was observed in the data for both groups. The analysis presented in the following sections aimed to explore the nature of this 'deviation' within the datasets.

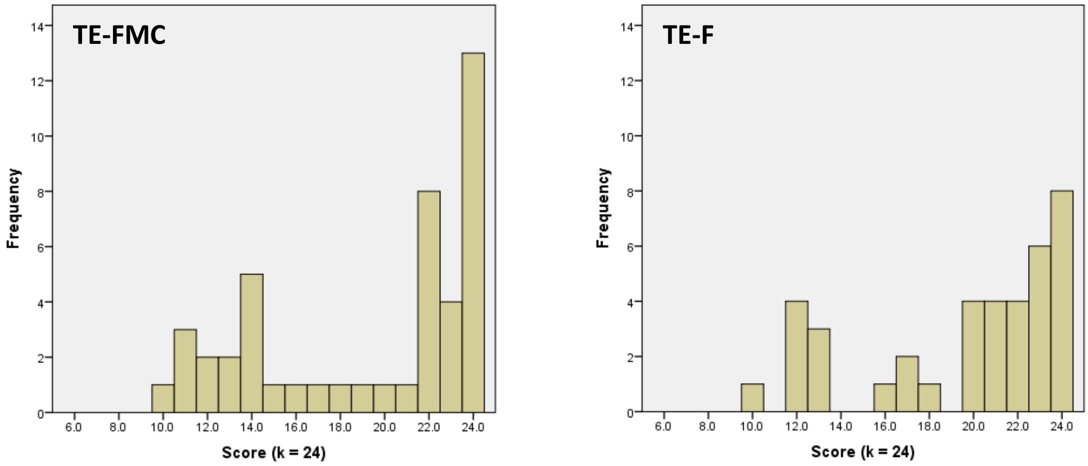
5.2.1 Distribution of scores

5.2.1.1 Examination of histograms

In order to look at the distribution of scores within each group and for each task, histograms of the learners' post-test scores were generated and examined (Figures 5.1a, 5.1b, 5.1c,).

[71] **TE-FMC**: Component 1, eigenvalue of 3.383 (56.38% variance). **TE-F**: Component 1, eigenvalue of 3.621 (60.36% variance)

Sentence Matching task



Gap-fill task

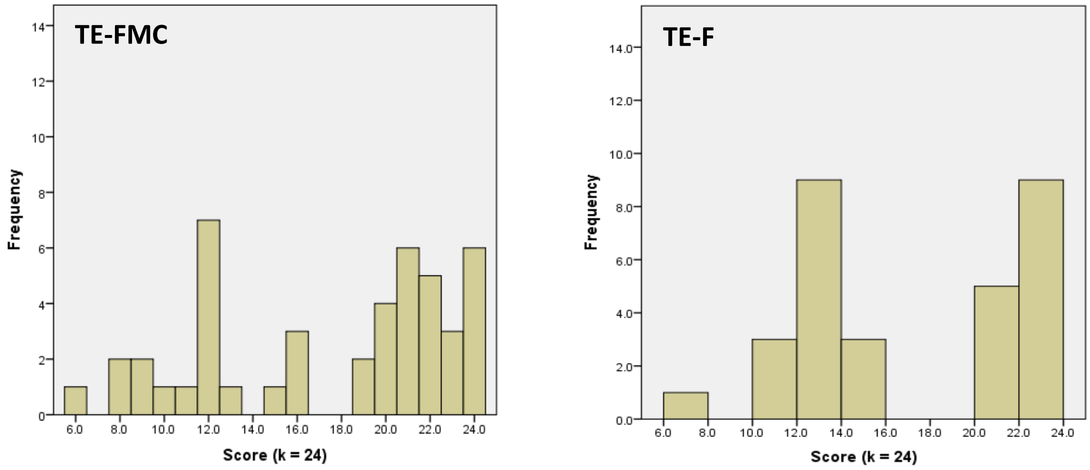


Figure 5.1a: Distribution of scores on the written tasks (post-test)

Sentence Reconstruction (Explanation) task

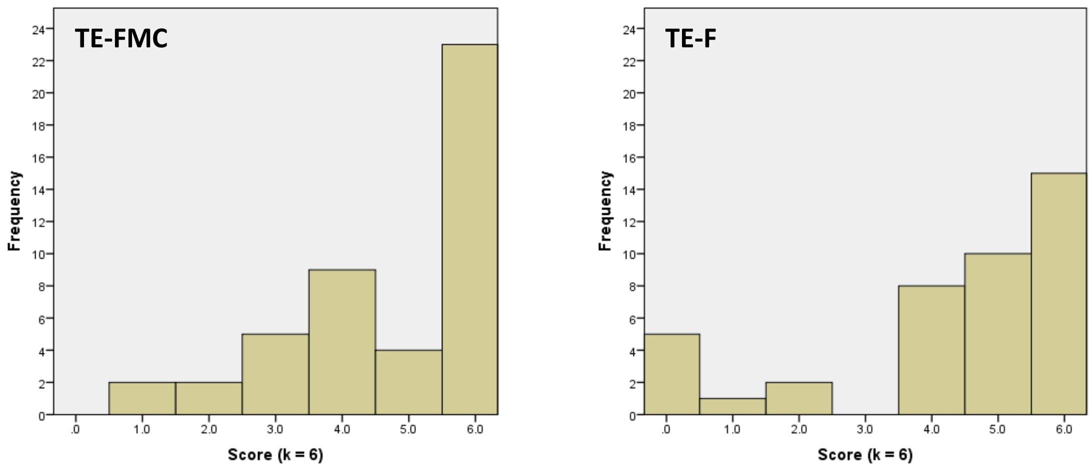
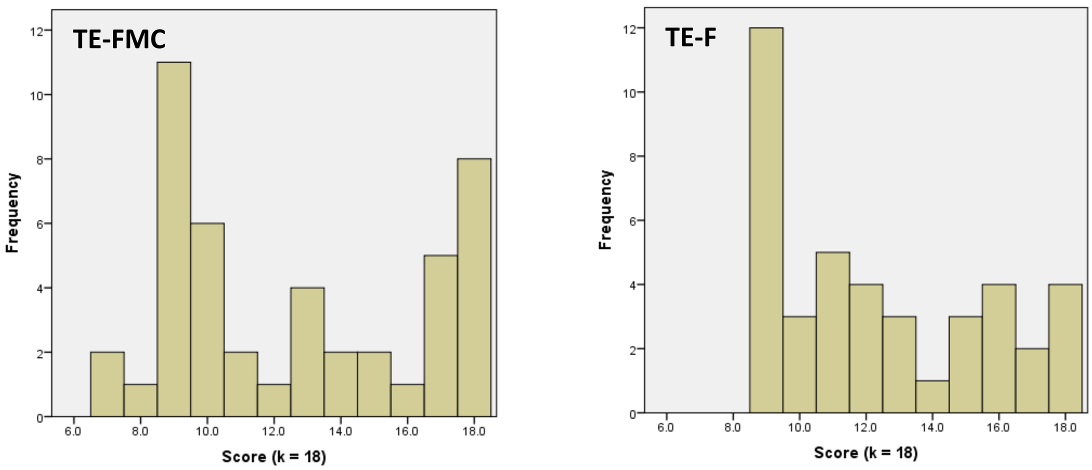
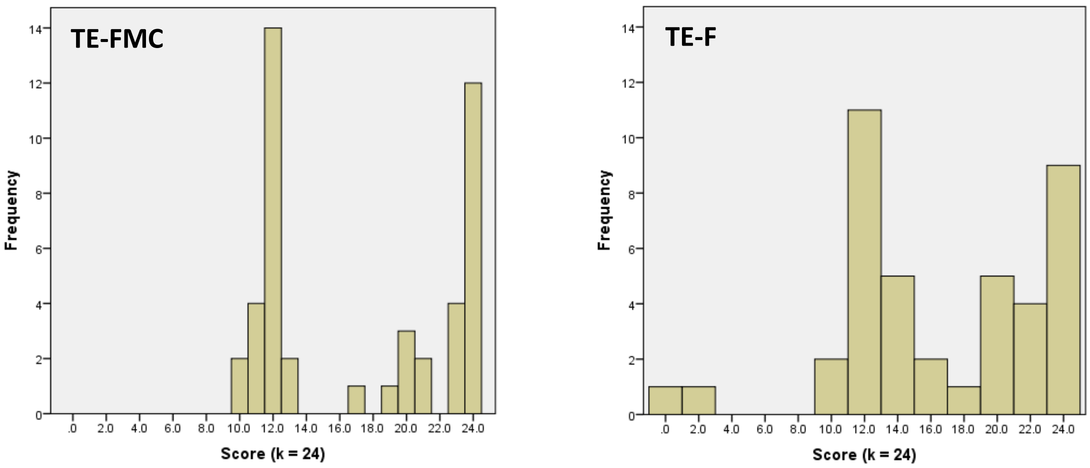


Figure 5.1b: Distribution of scores on the metalinguistic task (post-test)

Act-Out Comprehension task



Act-Out Production task



Sentence Repetition task

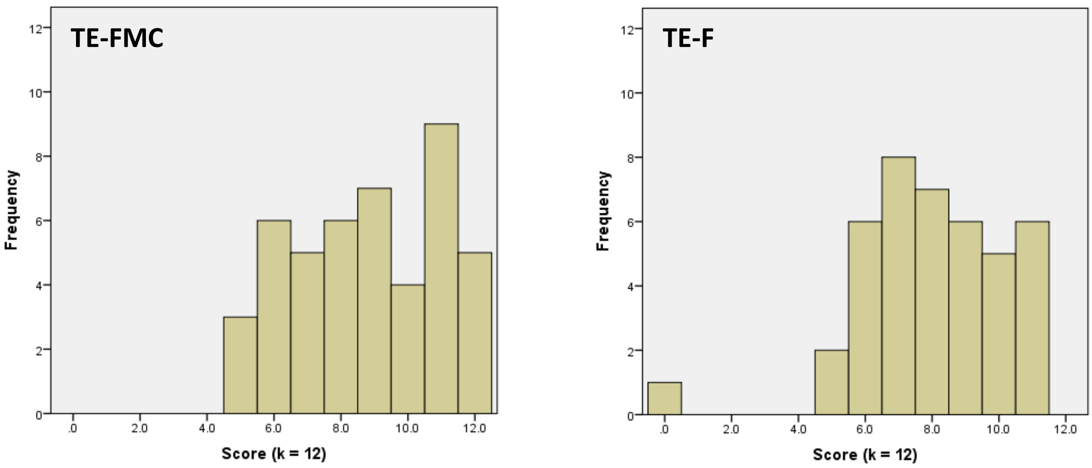


Figure 5.1c: Distribution of scores on the oral tasks (post-test)

Regarding the written tasks (Figure 5.1a), there was a distinctive bimodal distribution within the data for both tasks and both groups. Two clear peaks can be seen, the first at around 12 to 14 (chance level) and the second at 24 (highest score possible). This finding is notable given that there seems to be a clear divide within each group of learners; there are those learners who were performing at, or near, ceiling level (i.e. scoring 24) and those learners who seem to have made no improvement at post-test and continued to perform at chance level (i.e. scoring 12). A similar pattern was observed for the metalinguistic task (Figure 5.1b) and the oral tasks (Figure 5.1c), with three exceptions:

1) For the TE-FMC group on the Sentence Reconstruction (Explanation) task there was a negative skew within the dataset, with a just over half of the learners (*n* = 23 out of 45) scoring the maximum number of points available. For the TE-F group on the other hand whilst a number of learners (*n* = 15 out of 41) scored the maximum 6 points there was also a second smaller peak evident at 0 (baseline)
2) For the Sentence Repetition task both groups' scores are more spread. For the TE-FMC group the largest peak is close to ceiling level (*Mdn* = 11), whereas for the TE-F group the largest peak is close to chance level (*Mdn* = 7)
3) For the TE-F group on the Act-Out Comprehension task there was a large group of learners scoring 9 (out of 18), however there was no second peak at the top end of the scores

These observations indicated that there was a certain amount of variation between the two groups at post-test, with a larger number of TE-FMC learners receiving maximum scores than in the TE-F group.

As at post-test, for the delayed post-test data, a bimodal distribution was again found on each task for both groups (Appendix 20).

5.2.1.2 Identification of sub-groups

In order to investigate this variance in more detail, the learners within the TE-FMC and TE-F groups were separated into three sub-groups; 'Got It', 'Middle' and 'Not Got It'. A Middle group was included in order to allow for the fact that there may have been some learners who had not yet fully acquired the target grammatical feature and consequently were not able to reliably apply the new grammatical rule. Table 5.8 details the criteria by which each learner was allocated to a given sub-group for each task. It is important to note that the boundaries for each sub-group were identified intuitively, i.e. based on an examination of the visual pattern evident in the histograms for each task. Allocation was based on the learners' total score (*t*) on each task.

Table 5.8 Criteria for allocation to sub-group

Task	Max. Score	Sub-group		
		Not Got It	Middle	Got It
Sentence Matching	24	t < = 16	17 < t < 19	20 < = t
Gap-fill	24	t < = 16	17 < t < 19	20 < = t
Act-Out Comprehension	18	t < = 12	13 < t < 14	15 < = t
Act-Out Production	24	t < = 16	17 < t < 19	20 < = t
Sentence Repetition	12	t < = 8	t = 9	10 < = t
Sentence Reconstruction (E)	6	t < = 2	3 < t < 4	5 < = t

Sections 5.2.2 to 5.2.7 present the descriptive and inferential statistical analysis of:

a) the number of TE-FMC versus TE-F learners within the Got It and Not Got It sub-groups at post- and delayed post-test respectively[72] (between-group; Chi-square)
b) whether there was any change between post- and delayed post-test in the number of learners within each sub-group for the TE-FMC and TE-F groups respectively (within-group; McNemar)

5.2.2 *Sentence Matching* task

Table 5.9 presents the distribution of TE-FMC and TE-F participants within each sub-group. Only minimal differences were observed between the two groups.

Table 5.9: Distribution of learners within each sub-group (Sentence Matching)

Group	Sub-group	Post-test		Delayed post-test	
		N	%	*N*	%
TE-FMC (*N* = 45)	**Got It (> 20)**	27	60	28	62
	Middle (17-19)	3	7	2	4
	Not Got It (< 16)	15	33	15	33
TE-F (*N* = 38)	**Got It (> 20)**	26	68	24	63
	Middle (17-19)	3	8	1	3
	Not Got It (< 16)	9	24	13	34
Total (*N* = 83)	**Got It (> 20)**	53	64	52	63
	Middle (17-19)	6	7	3	3
	Not Got It (< 16)	24	29	28	34

Comparing between-group

The contingency tables for the post- and delayed post-test data are presented in Table 5.10.

[72] The aim of this analysis was to investigate any differences in the number of learners performing at ceiling versus baseline level between the TE-FMC and TE-F groups; therefore the Middle group was excluded.

Table 5.10: Contingency table for Chi-square analysis (Sentence Matching)

Time	Group	*N*	Got It		Not Got It	
			Observed	**Expected**	**Observed**	**Expected**
Post-test	**TE-FMC**	**42**	27	*28.9*	15	*13.1*
	TE-F	**35**	26	*24.1*	9	*10.9*
Delayed post-test	**TE-FMC**	**43**	28	*28*	15	*15.1*
	TE-F	**37**	24	*24.1*	13	*13*

Analysis via a two way (2x2) group independence Chi-square revealed no relationship between the TE-FMC and TE-F groups and the distribution of learners across the Got It and Not Got It sub-groups at either post- (χ^2 (1, *N* = 77) = .890, *p* = .35, *w* = .11) or delayed post-test (χ^2 (1, *N* = 80) = .001, *p* = .98, *w* = .003).

Comparing post- and delayed post-test

For the TE-FMC group (*N* = 45) there was relatively little movement between sub-groups from post- to delayed post-test; 21 learners were categorised as Got It at post-test and remained so at delayed post-test and nine learners remained in the Not Got It sub-group between post-test and delayed post-test. However it should be noted that six learners were found to move from the Got It to the Not Got It sub-group at delayed post-test, reflecting a substantial drop in the scores of these participants (Table 5.11). Nevertheless analysis found no change in the number of participants within each of the sub-groups between post- and delayed post-test test (McNemar, χ^2 (3, *N* = 45) = 5.400, *p* = .145).

Table 5.11: Crosstabulation of sub-group*time (Sentence Matching)

			Delayed post-test			
			Not Got It	**Middle**	**Got It**	**Total**
TE-FMC	**Post-test**	**Not Got It**	9	2	4	15
		Middle	0	0	3	3
		Got It	6	0	21	27
		Total	15	2	28	45
TE-F	**Post-test**	**Not Got It**	9	0	0	9
		Middle	3	0	0	3
		Got It	1	1	24	26
		Total	13	1	24	38

For the TE-F group (*N* = 38), as with the TE-FMC group, very little change was observed between post- and delayed post-test; only five participants were found to change sub-group (Table 5.11). Accordingly no difference was found between the two time points (McNemar, χ^2 (3, *N* = 38) = 5.000, *p* = .172).

Summary of sub-group analysis for Sentence Matching task

No difference was found between the TE-FMC and TE-F groups in terms of the distribution of learners within the Got It and Not Got It sub-groups; two thirds of the learners were categorised as Got It, confirming that they had made substantial improvement on the Sentence Matching task. Additionally no change in the distribution of learners across the sub-groups was observed between post- and delayed post-test. Nevertheless it is important to note that following their respective interventions one third of the participants in each group remained in the Not Got It sub-group at both time points, indicating that these participants had made no improvement between pre- and post-test (and delayed post-test) in their written comprehension of the target grammatical feature.

5.2.3 *Gap-Fill* task

Table 5.12 details the distribution of participants within each sub-group at post- and delayed post-test. At post-test the Got It sub-group contained more than half of the learners (TE-FMC, 56%; TE-F, 58%). At delayed post-test, however, a higher percentage of learners were categorised as Not Got It. Additionally the number of participants within the Middle sub-group had increased for both groups at delayed post-test.

Table 5.12: Distribution of learners within each sub-group (Gap-fill)

Group	Sub-group	Post-test		Delayed post-test	
		N	%	*N*	%
TE-FMC (*N* = 45)	**Got It (> 20)**	25	56	15	33
	Middle (17-19)	2	4	5	11
	Not Got It (< 16)	18	40	25	56
TE-F (*N* = 38)	**Got It (> 20)**	22	58	16	42
	Middle (17-19)	0	0	3	8
	Not Got It (< 16)	16	42	19	50
Total (*N* = 83)	**Got It (> 20)**	47	57	31	37
	Middle (17-19)	2	2	8	10
	Not Got It (< 16)	34	41	44	53

Comparing between-group

The contingency table for the post- and delayed post-test data (Table 5.13) shows that the observed count within both the TE-FMC and TE-F groups is the same as the expected count across both the Got It and Not Got It sub-groups. Accordingly the results of the analysis revealed no relationship between group membership (TE-FMC, TE-F) and the distribution of learners across the two sub-groups at post-test (2x2 Chi-square, χ^2 (1, *N* = 81) = .000, *p* = .982) or at delayed post-test (χ^2 (1, *N* = 75) = .519, *p* = .471).

Table 5.13: Contingency table for Chi-square analysis (Gap-fill)

Time	**Group**	***N***	**Got It**		**Not Got It**	
			Observed	**Expected**	**Observed**	**Expected**
Post-test	**TE-FMC**	**43**	25	*25*	18	*18*
	TE-F	**38**	22	*22*	16	*16*
Delayed post-test	**TE-FMC**	**40**	15	*16.5*	25	*23.5*
	TE-F	**35**	16	*14.5*	19	*20.5*

Comparing post- and delayed post-test

Considering the TE-FMC group first, although a number of learners remained within the Got It (n = 11) and Not Got It sub-groups (n = 13) between post- and delayed post-test, there were a substantial number of learners (n = 12) who moved from the Got It sub-group to the Not Got It sub-group at delayed post-test (Table 5.14). In contrast only three learners were found to move in the opposite direction (i.e. from Not Got It to Got it at delayed post-test). For the TE-FMC group the change in distribution of learners between post- and delayed post-test was approaching significance (McNemar, χ^2 (3, N = 45) = 7.733, p = .052), reflecting the observed movement of learners from the Got It to the Not Got It sub-group between post- and delayed post-test.

Table 5.14: Crosstabulation of sub-group*time (Gap-fill)

			Delayed post-test			
			Not Got It	**Middle**	**Got It**	**Total**
TE-FMC	**Post-test**	**Not Got It**	13	2	3	18
		Middle	0	1	1	2
		Got It	12	2	11	25
		Total	25	5	15	45
TE-F	**Post-test**	**Not Got It**	13	2	1	16
		Middle	-	-	-	-
		Got It	6	1	15	22
		Total	19	3	16	38

For the TE-F group, due to the fact that the Middle sub-group was ‘empty’ at post-test it was not possible to carry out analysis using the McNemar test. Nevertheless an examination of the crosstabulation revealed that a certain amount of movement was also evident in the TE-F group between post- and delayed post-test, with six learners moving from the Got It to the Not Got It sub-group.

Summary of sub-group analysis for Gap-fill task

There was no difference between the TE-FMC and TE-F groups in terms of the distribution of learners across the sub-groups at post- and delayed post-test respectively, but a certain amount of movement was observed for both groups between the two time points. A number of learners moved to the Not Got It sub-group at delayed post-test, reflecting a decrease in their performance on this task. It is important to note that no change was found in either the TE-FMC or TE-F learners' scores between post- and delayed post-test when analysed via Friedman's ANOVA (section 4.3.2.1). Nevertheless an examination of the median scores for each group indicated a more substantial change, with a decrease evident in both groups' median post-test scores (TE-FMC: post, *Mdn* = 20, delayed, *Mdn* = 15; TE-F: post, *Mdn* = 21, delayed, *Mdn* = 16.5) (see Table 4.4).

5.2.4 *Act-Out Comprehension* task

Table 5.15 details the distribution of learners within each sub-group. For both the TE-FMC and TE-F groups, the Not Got It sub-group contained a larger number of learners than the Got It sub-group. Further the descriptive statistics suggest a small increase in the number of learners within the Got It sub-group between post- and delayed post-test.

Table 5.15: Distribution of learners within each sub-group (Act-Out Comprehension)

Group	Sub-group	Post-test		Delayed post-test	
		N	%	*N*	%
TE-FMC (*N* = 45)	**Got It (> 20)**	16	36	19	42
	Middle (17-19)	6	13	4	9
	Not Got It (< 16)	23	51	22	49
TE-F (*N* = 41)	**Got It (> 20)**	13	32	16	39
	Middle (17-19)	4	10	3	7
	Not Got It (< 16)	24	59	22	54
Total (*N* = 86)	**Got It (> 20)**	29	34	35	41
	Middle (17-19)	10	11	7	8
	Not Got It (< 16)	47	55	44	51

Comparing between-group

Table 5.16 details the contingency tables for the Got It and Not Got it sub-groups at post- and delayed post-test. No relationship was found between group membership (TE-FMC or TE-F) and the count of learners in the two sub-groups at post-test (2x2 Chi-square, χ^2 (1, *N* = 76) = .279, *p* = .597) or delayed post-test (χ^2 (1, *N* = 79) = .143, *p* = .705).

Table 5.16: Contingency table for Chi-square analysis (Act-Out Comprehension)

Time	Group	*N*	Got It		Not Got It	
			Observed	Expected	Observed	Expected
Post-test	**TE-FMC**	**39**	16	*14.9*	23	*24.1*
	TE-F	**37**	13	*14.1*	24	*22.9*
Delayed post-test	**TE-FMC**	**41**	19	*18.2*	22	*22.8*
	TE-F	**38**	16	*16.8*	22	*21.2*

Comparing post- and delayed post-test

The crosstabulation of sub-group*time is presented in Table 5.17. For the TE-FMC group, a majority of the learners remained in the same sub-group between post- and delayed post-test (Not Got It, n = 17; Got It, n = 12). Accordingly analysis revealed no change in the distribution of the learners across the sub-groups between post- and delayed post-test (McNemar, χ^2 (3, N = 45) = 1.067, p = .785). For the TE-F group 18 of the learners remained in the Not Got It sub-group between post- and delayed post-test and only eight in the Got It sub-group. Five learners moved from the Not Got It to Got It sub-group between post- and delayed post-test and three moved in the opposite direction; nevertheless no difference was found between post- and delayed post-test (χ^2 (3, N = 41) = .700, p = .873).

Table 5.17: Crosstabulation of sub-group*time (Act-Out Comprehension)

			Delayed post-test			
			Not Got It	**Middle**	**Got It**	**Total**
TE-FMC	**Post-test**	**Not Got It**	17	2	4	23
		Middle	3	0	3	6
		Got It	2	2	12	16
		Total	22	4	19	45
TE-F	**Post-test**	**Not Got It**	18	1	5	24
		Middle	1	0	3	4
		Got It	3	2	8	13
		Total	22	3	16	41

Summary of sub-group analysis for Act-Out Comprehension task

Taken together the analysis of the Act-Out Comprehension task revealed that approximately half of the learners in both the TE-FMC and TE-F groups made minimal, or no, improvement on this task at post-test and were therefore classified as Not Got It at both the post- and delayed post-test. Additionally there was no change in the distribution of learners across the three sub-groups between post- and delayed post-test, with only a small number of learners in either group moving from the Not Got It to Got It sub-group. It is important to note that a significant improvement was observed in the analysis of both the TE-FMC and TE-F groups' scores on this task (section 4.4.1). The sub-group analysis suggests that this improvement can be accounted for by the ceiling level performance of those learners classified as Got it within the TE-FMC and TE-F groups at post-test.

5.2.5 *Act-Out Production* task

Table 5.18 (over page) presents the distribution of learners within each sub-group. At post-test a higher percentage of TE-FMC learners (51%) were categorised as Got It, than for the TE-F group (36%); however this difference had reduced by delayed post-test.

Comparing between-group

Table 5.19 details the contingency tables for the TE-FMC and TE-F groups at post- and delayed post-test. Although there were some slight differences between the expected and observed counts for each group, the analysis revealed no relationship between group membership (TE-FMC, TE-F) and the distribution of learners across the two sub-groups at post-test (2x2 Chi-square, χ^2 (1, N = 82) = .912, p = .339). Likewise at delayed post-test no relationship was observed (χ^2 (1, N = 86) = .297, p = .586).

Table 5.18: Distribution of learners within each sub-group (Act-Out Production)

Group	Sub-group	Post-test		Delayed post-test	
		N	**%**	***N***	**%**
TE-FMC (*N* = 45)	**Got It (> 20)**	23	51	23	51
	Middle (17-19)	0	0	3	7
	Not Got It (< 16)	22	49	19	42
TE-F (*N* = 41)	**Got It (> 20)**	15	36	20	49
	Middle (17-19)	4	10	0	0
	Not Got It (< 16)	22	54	21	51
Total (*N* = 86)	**Got It (> 20)**	38	44	43	50
	Middle (17-19)	4	5	3	3
	Not Got It (< 16)	44	51	40	47

Table 5.19: Contingency table for Chi-square analysis (Act-Out Production)

Time	Group	*N*	Got It		Not Got It	
			Observed	**Expected**	**Observed**	**Expected**
Post-test	**TE-FMC**	**45**	23	*20.9*	22	*24.1*
	TE-F	**37**	15	*17.1*	22	*19.9*
Delayed post-test	**TE-FMC**	**42**	23	*21.8*	19	*20.2*
	TE-F	**41**	20	*21.2*	24	*19.8*

Comparing post- and delayed post-test

The crosstabulation of sub-group*time is presented in Table 5.20. Due to the fact that there were no TE-FMC group participants within the Middle sub-group at post-test, and no TE-F group participants in the Middle sub-group at delayed post-test, it was not possible to conduct the McNemar test on this data. Nevertheless an examination of the crosstabulation indicated that there was little movement between the sub-groups, with a majority of learners from both groups remaining in the Not Got It or Got It sub-groups between post- and delayed post-test.

Table 5.20: Crosstabulation of sub-group*time (Act-Out Production)

			Delayed post-test			
			Not Got It	**Middle**	**Got It**	**Total**
TE-FMC	**Post-test**	**Not Got It**	18	1	3	22
		Middle	-	-	-	-
		Got It	1	2	20	23
		Total	19	3	23	45
TE-F	**Post-test**	**Not Got It**	18	-	4	22
		Middle	1	-	3	4
		Got It	2	-	13	15
		Total	21	-	20	41

Summary of sub-group analysis for Act-Out Production task

The examination of the sub-groups for the Act-Out Production task suggested that for both the TE-FMC and TE-F groups a substantial number of learners were performing at a high level on this task and were therefore classified as and remained within the Got It sub-group at post- and delayed post-test. However it should be noted that there were also a number of participants who continued to perform at baseline on this task at post- and delayed post-test and therefore demonstrated minimal improvement in their production of the target feature.

5.2.6 *Sentence Repetition* task

The descriptive statistics (Table 5.21) suggest a greater level of divergence between the TE-FMC and TE-F groups, than was seen on the other activities.

Table 5.21: Distribution of learners within each sub-group (Sentence Repetition)

Group	**Sub-group**	**Post-test**		**Delayed post-test**	
		N	**%**	***N***	**%**
TE-FMC (*N* = 45)	**Got It (> 10)**	18	40	23	51
	Middle (9)	7	16	5	11
	Not Got It (< 8)	20	44	17	38
TE-F (*N* = 41)	**Got It (> 10)**	11	27	12	29
	Middle (9)	6	15	5	12
	Not Got It (< 8)	24	58	24	59
Total (*N* = 86)	**Got It (> 10)**	29	34	35	41
	Middle (9)	13	15	10	11
	Not Got It (< 8)	44	51	41	48

At post-test the Got It sub-group contained 40% of the TE-FMC group compared to only 27% of the TE-F group. Additionally the percentage of TE-FMC learners' in the Got It sub-group rose to 51% at delayed post-test, whereas only a minimal change was observed in the TE-F group (29% at delayed post-test). In contrast the Not Got It sub-group contained more than half of the TE-F learners at post- (58%) and delayed post-test (59%), compared to only 44% of the TE-FMC learners at post-test, which had decreased to 38% by delayed post-test.

Comparing between-group

An initial examination of the contingency table (Table 5.22) reflected the observations made based on the descriptive statistics. For the TEFMC group, the observed count of participants within the Got It sub-group was higher than the expected count, whereas the observed count for the Not Got It sub-group was lower than the expected count. The opposite pattern was found for the TE-F group. Nevertheless, no relationship was found between group (TE-FMC, TE-F) and the count of participants within the two sub-groups at post-test (2x2 Chi-square, χ^2 (1, N = 73) = 1.933, p = .164, w = .163). For the delayed post-test data, however, a significant relationship was revealed between the TE-FMC and TE-F groups and the distribution of learners across the Got It and Not Got It sub-groups (χ^2 (1, N = 76) = 4.454, p = .035, w = .242).

Table 5.22: Contingency table for Chi-square analysis (Sentence Repetition)

			Got It		Not Got It	
Time	**Group**	***N***	**Observed**	**Expected**	**Observed**	**Expected**
Post-test	**TE-FMC**	**38**	18	*15.1*	20	*22.9*
	TE-F	**35**	11	*13.9*	24	*21.1*
Delayed post-test	**TE-FMC**	**40**	23	*18.4*	17	*21.6*
	TE-F	**36**	12	*16.6*	24	*19.4*

Comparing post- and delayed post-test

For the TE-FMC group an equal number of learners remained in the Got It and Not Got It sub-groups between post- and delayed post-test (n = 15 each). In total eight learners (three Not Got It, five Middle) moved to the Got It sub-group at delayed post-test, reflecting an improvement in their performance:

Table 5.23: Crosstabulation of sub-group*time (Sentence Repetition)

			Delayed post-test			
			Not Got It	**Middle**	**Got It**	**Total**
TE-FMC	**Post-test**	**Not Got It**	15	2	3	20
		Middle	1	1	5	7
		Got It	1	2	15	18
		Total	17	5	23	45
TE-F	**Post-test**	**Not Got It**	19	2	3	24
		Middle	2	2	2	6
		Got It	3	1	7	11
		Total	24	5	12	41

For the TE-F group, however, a higher number of learners remained in the Not Got It sub-group (n = 19) compared to the Got It sub-group (n = 7) at delayed post-test. Some movement was observed with five learners (three Not Got It, two Middle) moving to the Got It sub-group and five learners (three Got It, two Middle) moving to the Not Got It sub-group (Table 5.23). Despite the small amount of movement in both groups, analysis found no change in the distribution of learners across the three sub-groups, for either the TE-FMC (McNemar, χ^2 (1, N = 45) = 2.169, p = .454) or TE-F group (χ^2 (1, N = 41) = .333, p = .954).

Summary of sub-group analysis for Sentence Repetition task

Examination of the count of learners across the three sub-groups on this task yielded a different pattern of results to the other outcome measures. Firstly the findings confirmed that there were a significantly higher number of TE-FMC participants in the Got It sub-group than for the TE-F group at delayed post-test. Additionally a higher number of TE-F participants were categorised as Not Got It than for the TE-FMC group at both post- and delayed post-test. This divergence between the TE-FMC and TE-F groups is reflective of the approaching significant difference observed between the two groups' scores on this task at delayed post-test (section 4.4.3.1).

5.2.7 *Sentence Reconstruction (Explanation)* task

At post-test, 60% of both the TE-FMC and TE-F groups were categorised as Got It on the Sentence Reconstruction (Explanation) task, compared to only 9% (TE-FMC) and 20% (TE-F) in the Not Got It sub-group (Table 5.24). However the distribution of learners across the three sub-groups changed substantially at delayed post-test, with a decrease in the percentage of learners in the Got It sub-group and an increase in the Not Got It sub-group.

Table 5.24: Distribution of learners within each sub-group (Sentence Reconstruction-Explanation)

Group	Sub-group	Post-test		Delayed post-test	
		N	**%**	***N***	**%**
TE-FMC (*N* = 45)	**Got It (> 5)**	27	60	19	42
	Middle (3-4)	14	31	11	25
	Not Got It (< 2)	4	9	15	33
TE-F (*N* = 41)	**Got It (> 5)**	25	60	8	20
	Middle (3-4)	8	20	16	39
	Not Got It (< 2)	8	20	17	41
Total (*N* = 86)	**Got It (> 10)**	52	60	27	31
	Middle (9)	22	26	27	11
	Not Got It (< 8)	12	14	32	37

Comparing between-group

Table 5.25 details the contingency table for the post- and delayed post-test data. Analysis found no relationship between group (TE-FMC, TE-F) and the count of learners within the sub-groups at post-test (2x2 Chi-square, χ^2 (1, N = 64) = 1.349, p = .245, w = .145). At delayed post-test, however, the results of the 2x2 Chi-square test were approaching significance (χ^2 (1, N = 59) = 3.311, p = .069, w = .237), reflecting a divergence between the TE-FMC and TE-F groups in the number of learners within the Got it sub-group at delayed post-test.

Table 5.25: Contingency table for Chi-square analysis (Sentence Reconstruction - Explanation)

Time	Group	*N*	Got It		Not Got It	
			Observed	**Expected**	**Observed**	**Expected**
Post-test	**TE-FMC**	**31**	27	*25.2*	4	*5.8*
	TE-F	**33**	25	*26.8*	8	*6.2*
Delayed post-test	**TE-FMC**	**34**	19	*15.6*	15	*18.4*
	TE-F	**25**	8	*11.4*	17	*13.6*

Comparing post- and delayed post-test

The crosstabulation of sub-groups and time points is presented in Table 5.26. For the TE-FMC group a substantial amount of movement was observed between post- and delayed post-test, with 14 learners (eight Got It, six Middle) moving to the Not Got It sub-group at delayed post-test. Correspondingly the results of the McNemar test were approaching significance (χ^2 (1, N = 45) = 7.571, p = .056). With regard to the TE-F group, an examination of the crosstabulation (Table 5.26) revealed a large amount of movement with 11 learners (eight Got It, three Middle) moving to the Not Got It sub-group at delayed post-test, as well as ten learners moving to the Middle sub-group. Accordingly the results of the analysis were significant (McNemar, χ^2 (1, N = 41) = 15.564, p = .001). These findings suggested that for both groups, in particular the TE-F group, there was a change in the distribution of learners across the three sub-groups between post- and delayed post-test.

Table 5.26: Crosstabulation of sub-group*time (Sentence Reconstruction - Explanation)

			Delayed post-test			
			Not Got It	**Middle**	**Got It**	**Total**
TE-FMC	**Post-test**	**Not Got It**	1	1	2	4
		Middle	6	4	4	14
		Got It	8	6	13	27
		Total	15	11	19	45
TE-F	**Post-test**	**Not Got It**	6	2	0	8
		Middle	3	4	1	8
		Got It	8	10	7	25
		Total	17	16	8	41

Summary of sub-group analysis for Sentence Reconstruction task

Analysis revealed a different pattern to that of the other outcome measures; for this task the number of learners performing to a high level at delayed post-test decreased, with fewer participants in both the TE-FMC and TE-F groups being categorised as Got It. Additionally there was a significant change (approaching significance for the TE-FMC group) in the distribution of learners across the three sub-groups, with a large number of learners moving down to the Not Got It sub-group at delayed post-test. This finding is line with the significant decrease observed in the TE-FMC and TE-F groups' scores at delayed post-test (section 4.6.1.2).

5.2.8 Summary of findings for sub-group analysis

The analysis presented in sections 5.2.2 to 5.2.7 aimed to investigate the large standard deviation found in both the TE-FMC and TE-F groups' data at post- and delayed post-test. An examination of the histograms for each task revealed a striking bimodal distribution across the activities. Further the sub-division of participants into sub-groups (Got It, Middle, Not Got It) revealed that in both groups, across a majority of the outcome measures, the participants either tended to be performing at ceiling level (Got It) or at baseline level (Not Got It) at post- and delayed post-test.

Across the written tasks and the Act-Out Comprehension and Production tasks there were no differences between the TE-FMC and TE-F groups in terms of the distribution of learners across the sub-groups. Additionally, there was minimal movement of participants between the three sub-groups from post- to delayed post-test. There were two exceptions to these findings; firstly on the Sentence Repetition task a higher number of TE-FMC learners were categorised at Got it than for the TE-F group at delayed post-test. Secondly for the Sentence Reconstruction (Explanation) task a significant change was observed in the distribution of learners across the three sub-groups with a substantial decrease in the number of learners within the Got It sub-group and an increase in the Not Got It sub-group at delayed post-test (for both the TE-FMC and TE-F groups). This finding is reflective of the significant drop in scores observed on this task.

Taken together these findings suggest that, despite the significant improvement observed in both the TE-FMC and TE-F groups' performance across the six outcome measures, this improvement can be accounted for by the ceiling level performance of a sub-group of learners within each of the TE-FMC and TE-F groups (see section 6.4.4 for discussion).

5.3 Role of grammatical sensitivity

The findings presented in section 5.2 highlight an important consideration in language teaching research, as to the way in which individuals respond to a particular teaching approach and the significance of individual learner differences. The following section will therefore examine the participants' performance on the grammatical sensitivity task, in order to investigate the role of grammatical sensitivity in mediating the TE-FMC and TE-F groups' performance across the six outcome measures following their respective teaching interventions[73].

5.3.1 Performance on the grammatical sensitivity task

The descriptive statistics for the grammatical sensitivity task are presented in Table 5.27; an initial examination suggested that the TE-FMC and TE-F groups were performing at an equivalent level.

Table 5.27: Descriptive statistics for the grammatical sensitivity task

Group	N	*M*	*SD*	*Mdn*
TE-FMC	41	20.27	4.62	22
TE-F	40	20.38	5.51	22
Total	81	20.32	5.05	22

Before carrying out further analysis, the learners' scores were analysed in order to assess the normality and homogeneity of variance within the dataset. The Shapiro-Wilk test of normality revealed that both the TE-FMC ($W(41) = 0.863$, $p = .001$) and TE-F ($W(37) = 0.837$, $p = .001$) datasets were non-normally distributed. Accordingly the non-parametric Mann Whitney U test was utilised and in line with the initial observation, no difference was found between the TE-FMC and TE-F groups ($U = 872.000$, $z = .493$, $p = .622$).

5.3.2 Grammatical sensitivity as a covariate

Previous research has demonstrated that aptitude, and in particular grammatical sensitivity, may be an intervening factor in the language acquisition process. Therefore an analysis of covariance (ANCOVA) was conducted on the data for each outcome measure in order to investigate the relationship between the learners' level of grammatical sensitivity, as measured on the grammatical sensitivity task, and their performance on the respective outcome measures following the intervention. Additionally it was hypothesised that this

[73] The Control group was excluded given that they did not receive any teaching intervention and there was no change in their performance between pre- and post-test.

analysis might shed light on the divergence observed on each of the tasks, between those learners who were categorised as 'Got It' at post- and delayed post-test and were performing at ceiling level, and those learners who were categorised as 'Not Got It' and continued to perform at baseline level despite the instructional treatment received (section 5.2). An ANCOVA was therefore utilised including the independent variables Group (two levels: TE-FMC, TE-F) and Time (three levels: pre-, post-, delayed post-test) and the learners' scores on the grammatical sensitivity task were introduced into the analysis as the covariate. It is important to acknowledge at this point, however, that given the non-normal distribution of the datasets, the results of the ANCOVA should be interpreted with caution.

Overall, the learners' scores on the grammatical sensitivity task were found to have a significant effect on the dependent variable (i.e. performance on the outcome measures) (Table 5.28). Examining the results of the six tasks, the covariate (grammatical sensitivity) was found to exert the largest influence on the Sentence Matching (written, comprehension) task, whereas the analysis of the Act-Out Comprehension and Act-Out Production tasks yielded the smallest effects of the covariate.

Table 5.28: ANCOVA results for effect of covariate (grammatical sensitivity task score)

Task		df	F	*p*	ηp^2	Power
Written	**S. Matching**	1,74	20.542	*0.001*	0.217	0.994
	Gap-fill	1,74	11.330	*0.001*	0.133	0.913
Oral	**A-O Comprehension**	1,76	6.373	*0.014*	0.077	0.703
	AO Production	1,76	7.609	*0.007*	0.091	0.777
	S. Repetition	1,76	17.843	*0.001*	0.190	0.986
Metalinguistic	**S. Reconstruction (E)**	1,76	17.319	*0.001*	0.186	0.984

It is important to note, however, that even after taking account of the effect of the covariate, a significant difference was still observed between the learners' pre- and post-test and pre- and delayed post-test performance on all of the tasks.

Summary of ANCOVA results

Grammatical sensitivity was a significant covariate for all of the outcome measures, although exerted only a marginal effect on some of the tasks. Nevertheless a significant change in the learners' scores on each outcome measure was still observed between the time points. Grammatical sensitivity, therefore, appears to have been one factor which affected the participants' performance when completing the outcome measures at post- and delayed post-test.

5.3.3 Relationship between grammatical sensitivity and performance on the outcome measures

In order to investigate in more detail the nature of the relationship between the participants' grammatical sensitivity and their performance on the respective outcome measures at post- and delayed post-test, a standard regression analysis was carried out for each outcome measure. Previous analysis (section 5.2) identified a substantial amount of variation (in the form of standard deviation) within the learners' scores on the outcome measures at post- and delayed post-test, characterised by the striking bi-modal distribution of scores (Got It versus Not Got It). Consequently the aim of the regression analysis was to investigate how much of this variation could be accounted for by the learners' level of grammatical sensitivity.

5.3.3.1 Correlation analysis

Table 5.29 (over page) summarises the results of Pearson's correlation between the learners' grammatical sensitivity scores and their scores on each of the tasks at post- and delayed post-test. Significant positive correlations were found between the grammatical sensitivity task and both groups' performance on the two written measures at post- and delayed post-test (Table 5.29), although for the TE-F group the correlation between the grammatical sensitivity and Gap-fill tasks was only approaching significance (r = .250, p = .068) at delayed post-test.

The results of the correlation for the oral tasks, however, revealed a different pattern to that of the written tasks (Table 5.29). A significant correlation was found between grammatical sensitivity and the Sentence Repetition task at post- and delayed post-test for both groups. However for the TE-FMC group no relationship was observed between the learners' grammatical sensitivity and their scores on either the Act-Out Comprehension or Act-Out Production task, at post- or delayed post-test. In contrast for the TE-F group a significant correlation was observed for all of the oral tasks, with the exception of the Act-Out Comprehension task at delayed post-test, for which the correlation was approaching significance (r = .250, p = .060). For the metalinguistic Sentence Reconstruction (Explanation) task, no correlation was observed for the TE-FMC group at post-test (r = -.013, p = .469), however a significant correlation was observed at delayed post-test. The TE-F groups' scores on this task were found to correlate with their grammatical sensitivity task scores at both post- and delayed post-test.

Table 5.29: Results of Pearson's correlation between grammatical sensitivity task scores and outcome measure scores

Task		Time	Group	N	*r*	*p*
Written	**Sentence Matching**	**Post**	TE-FMC	41	0.452	*0.001*
			TE-F	37	0.460	*0.002*
		Delayed post	TE-FMC	41	0.516	*0.000*
			TE-F	37	0.335	*0.021*
	Gap-fill	**Post**	TE-FMC	41	0.289	*0.033*
			TE-F	37	0.355	*0.016*
		Delayed post	TE-FMC	41	0.366	*0.009*
			TE-F	37	0.250	*0.068*
Oral	**Act-Out Comprehension**	**Post**	TE-FMC	40	0.218	*0.089*
			TE-F	40	0.290	*0.035*
		Delayed post	TE-FMC	40	0.207	*0.100*
			TE-F	40	0.250	*0.060*
	Act-Out Production	**Post**	TE-FMC	40	0.113	*0.244*
			TE-F	40	0.466	*0.001*
		Delayed post	TE-FMC	40	0.229	*0.077*
			TE-F	40	0.309	*0.026*
	Sentence Repetition	**Post**	TE-FMC	40	0.366	*0.010*
			TE-F	40	0.413	*0.004*
		Delayed post	TE-FMC	40	0.349	*0.014*
			TE-F	40	0.384	*0.007*
Metalinguistic	**Sentence Reconstruction (Explanation)**	**Post**	TE-FMC	40	-0.013	*0.469*
			TE-F	40	0.425	*0.003*
		Delayed post	TE-FMC	40	0.365	*0.010*
			TE-F	40	0.495	*0.001*

Summary of the correlation analysis

The examination of correlation coefficients suggested that there was a relationship between grammatical sensitivity and performance on the written tasks for both the TE-FMC and TE-F groups. In contrast the results for the oral tasks were more mixed. For the TE-FMC group the only significant correlation observed was for the Sentence Repetition task, whereas for the TE-F group significant correlations were more consistently observed across the three oral tasks. The metalinguistic task scores were found to correlate with the grammatical sensitivity task scores, except for the TE-FMC group at post-test.

5.3.3.2 Regression analysis

A summary of the regression analysis for each task is presented in Table 5.30 (over page). It is important to note that overall the r^2 values generated are relatively small (< .3), suggesting that grammatical sensitivity accounted for a relatively small amount of variance on any one task. Nevertheless the results indicated that grammatical sensitivity did have a statistically significant relationship with the learners' performance on at least some of the tasks.

Table 5.30: Results of standard regression analysis (predictor variable: grammatical sensitivity task scores)

Task		Time	Group	*N*	r^2	B	Intercept
Written	**Sentence Matching**	**Post**	TE-FMC	*41*	0.204**	0.494	9.054
			TE-F	*37*	0.212**	0.428	10.525
		Delayed post	TE-FMC	*41*	0.267**	0.527	8.023
			TE-F	*37*	0.113*	0.340	11.737
	Gap-fill	**Post**	TE-FMC	*41*	0.084	0.346	10.602
			TE-F	*37*	0.126*	0.403	9.546
		Delayed post	TE-FMC	*41*	0.134*	0.395	9.009
			TE-F	*37*	0.062	0.270	11.677
Oral	**Act-Out Comprehension**	**Post**	TE-FMC	*40*	0.047	0.171	9.219
			TE-F	*40*	0.084	0.169	8.899
		Delayed post	TE-FMC	*40*	0.043	0.165	9.493
			TE-F	*40*	0.062	0.146	9.258
	Act-Out Production	**Post**	TE-FMC	*40*	0.013	0.139	14.566
			TE-F	*40*	0.217**	0.523	5.434
		Delayed post	TE-FMC	*40*	0.053	0.276	12.800
			TE-F	*40*	0.096	0.366	9.340
	Sentence Repetition	**Post**	TE-FMC	*40*	0.134*	0.173	5.205
			TE-F	*40*	0.170**	0.167	4.597
		Delayed post	TE-FMC	*40*	0.122*	0.177	5.565
			TE-F	*40*	0.147*	0.152	5.060
Metalinguistic	**Sentence Reconstruction (Explanation)**	**Post**	TE-FMC	*40*	0.000	-0.004	4.834
			TE-F	*40*	0.180**	0.156	1.097
		Delayed post	TE-FMC	*40*	0.133*	0.172	-0.136
			TE-F	*40*	0.245**	0.167	-0.512

Overall the largest r^2 values were generated for the written tasks. For instance grammatical sensitivity was found to account for 20.4% (r^2 = .204) of the variance in the TE-FMC group's scores on the Sentence Matching task at post-test, and 26.7% (r^2 = .267) at delayed post-test. Similarly for the TE-F group 21.2% (r^2 = .212) of the variance was accounted for on the Sentence Matching task at post-test. For the Gap-Fill task, however, the results of the regression were more mixed, with grammatical sensitivity accounting for a significant amount of variance for the TE-F group at post-test (r^2 = .126), and the TE-FMC group at delayed post-test (r^2 = .134) only.

With regard to the oral tasks, the only task for which grammatical sensitivity was found to be a statistically significant explanatory variable was the Sentence Repetition task with 13.4% (r^2 = .134) and 12.2% (r^2 = .122) of the variance being explained for the TE-FMC and TE-F groups respectively at post-test (delayed post-test: TE-FMC, r^2 = .170; TE-F, r^2 = .147). In contrast grammatical sensitivity was not found to make a statistically significant contribution to the Act-Out Comprehension task scores for either group or to the Act-Out Production task scores for the TE-FMC group. It was found to explain 21.7% (r^2 = .217) of the variance in the TE-F group's scores on the Act-Out Production task at post-test. Additionally, with regard to the Sentence Reconstruction task, grammatical sensitivity was found to explain a significant amount of variance in the TE-F group's scores at post- and delayed post-test, as well as the TE-FMC group's scores at delayed post-test.

Summary of the regression analysis

The results of the regression analysis revealed several notable findings relating to the influence of grammatical sensitivity on the learners' performance at post- and delayed post-test. Firstly despite the fact that grammatical sensitivity was previously found to be a significant covariate across all of the outcome measures (section 5.3.2), the only two tasks for which the regression analysis yielded grammatical sensitivity as an explanatory variable for both groups at both time points were the Sentence Matching (written, comprehension) and Sentence Repetition (oral, production) tasks. Secondly at post-test grammatical sensitivity was consistently found to explain a larger percentage of the variance within the TE-F group's data than the TE-FMC group's data (for all tasks); this pattern was reversed at delayed post-test for the two written tasks.

Taken together, the findings of the regression analysis suggest that there was a relationship between grammatical sensitivity and certain measures used in this study. Nevertheless the small r^2 values yielded suggest that there were likely to be other factors, which also contributed to the learners' performance following the respective interventions.

5.4 Summary of Results 2

The focus of the second results chapter was on:

a) investigating the type of knowledge elicited by the six outcome measures following the TE-FMC and TE-F interventions
b) analysing the participants' performance on a more fine-grained individual level
c) analysing the role of grammatical sensitivity in mediating the learners' performance on the outcome measures at post- and delayed post-test

PCA revealed that at post-test the oral tasks were loading onto a separate component to the written and metalinguistic tasks. This finding indicated that the tasks may have been tapping into different types of knowledge; namely more explicit knowledge for the written and metalinguistic tasks and possibly a more implicit form of knowledge for the oral tasks. This pattern was also observed when the TE-FMC group data was analysed separately. For the TE-F group data, however, at post-test the oral and written tasks were found to load together and the metalinguistic task separately, suggesting that the TE-F group may have been relying on perhaps more implicit knowledge when completing both the written and oral tasks at post-test. In contrast at delayed post-test, all of the tasks were found to load together onto one component. Taken together with the significant decrease in the learners' scores on the metalinguistic task, this finding suggests that the tasks may have been eliciting a more implicit type of knowledge at delayed post-test, or, at least, that the learners' ability to articulate their explicit knowledge accurately had decreased (see section 6.3 for discussion).

The breakdown of learners into sub-groups (Got It, Middle, Not Got It) revealed that whilst a majority of the TE-FMC and TE-F learners tended to be performing at ceiling level (Got It) across the six outcome measures, there were a substantial number of learners who continued to perform at baseline level (Not Got It) at post- and delayed post-test. Therefore the overall significant improvement observed on each task at post-test can be attributed to the ceiling-level performance of a sub-group of learners within the TE-FMC and TE-F groups (see section 6.4.4 for discussion).

Finally the results of the ANCOVA indicated that grammatical sensitivity had a significant effect on the leaners' performance across the six outcome measures, although this effect was marginal for some of the tasks. The results of the correlation revealed a relationship between the learners' performance on the grammatical sensitivity task and their performance on the written tasks. The findings were more mixed for the remaining tasks. The regression analyses indicated that, as an explanatory variable, grammatical sensitivity accounted for a significant amount of the variance within the Sentence Matching task datasets for both groups and for the Gap-fill task to a certain extent. This finding was also observed for the oral Sentence Repetition task for both groups at both time points; however was not consistently observed in the other oral tasks or the metalinguistic task. Further the findings indicated that grammatical sensitivity may have played a larger role for the TE-F group than the TE-FMC group across all of the tasks, particularly at post-test (see section 6.4.5 for discussion).

Chapter 6: Discussion

This chapter will discuss key findings of the present study in relation to previous effect of instruction research. The learners' overall performance across the outcome measures will be discussed. The findings will then be considered in relation to the three research questions addressed in this study:

1. Does explicit grammar instruction improve young learners' a) comprehension and b) production of the target grammatical feature?

2. To what extent does explicit grammar instruction develop different types of knowledge of the target grammatical feature?

3. Following explicit information, is intentional practice in attending to the target form-meaning connection more beneficial than intentional practice in attending to the target grammatical form only?

6.1 Overall performance across the outcome measures

6.1.1 General findings

Examination of the results (Chapters 4 and 5) highlights a number of general findings. First, analysis of the pre-test performance of participants revealed that all three groups – task-essential form-meaning connection (TE-FMC), task-essential form-only (TE-F), Control – were performing at an equivalent, chance (or baseline) level prior to the intervention across the six outcome measures (written, oral and metalinguistic). At post-test, however, both the TE-FMC and TE-F groups were significantly outperforming the Control group across all tasks. In addition, no difference was found between the TE-FMC and TE-F groups at post- or delayed post-test on any of the tasks. The comparative performance of the three groups can, therefore, be expressed as follows:

(Pre-test) TE-FMC = TE-F = Control

(Post-test) TE-FMC = TE-F > Control

(Delayed post-test) TE-FMC = TE-F

With regard to the learners' performance over Time, no change in the Control group's performance occurred between pre- and post-test on any of the tasks. In contrast, there was significant improvement in both the TE-FMC and TE-F groups' performance across the outcome measures at post-test. Furthermore, both the TE-FMC and TE-F groups sustained their improved performance across the delayed post-test (nine weeks after the intervention). The one exception was the metalinguistic task (Sentence Reconstruction - Explanation), in which both the TE-FMC and TE-F learners' scores decreased at delayed post-test[74]. As such the respective performances of the three groups over Time can be summarised as follows:

(TE-FMC)	Pre < Post = Delayed Pre < Post > Delayed (*Sentence Reconstruction task*)
(TE-F)	Pre < Post = Delayed Pre < Post > Delayed (*Sentence Reconstruction task*)
(Control)	Pre = Post

6.1.2 Superiority of the interventions

The pattern of results detailed above occurred consistently across the five written and oral tasks, suggesting that the TE-FMC and TE-F interventions had an equivalent effect on both the learners' (written and oral) comprehension and their production of the target feature. The overall improvement in the TE-FMC and TE-F groups' performance at post-test is in line with previous research (for reviews see Norris & Ortega, 2000; Spada & Tomita, 2010) which has demonstrated that explicit instruction can have a beneficial effect on the learning of key grammatical features. In the present study the magnitude of change from pre- to post-test was large for both the TE-FMC (d = 1.63) and TE-F (d = 1.60) treatments. This is in line with the large mean effect size calculated for explicit FonF treatments (d = 2.08) in Norris and Ortega's (2000) meta-analysis as well as in Spada and Tomita's (2010) meta-analysis (d = 0.88 for simple forms, d = 0.84 for complex forms).

Crucially, in the present study, the improvement of the TE-FMC and TE-F groups was not reflected in the performance of the Control group at post-test on any of the tasks. Notably, a small mean effect size (d = 0.20) was yielded from the pre- and post-test Control group data, which is similar to that observed by Norris and Ortega (2000) (d = 0.30) and Spada and Tomita (2010) (d = 0.28) in their reviews of previous studies which had included a control group. Any change in a non-active control group can potentially illuminate the extent to which factors, such as practice, exposure and maturation, may have contributed to any improvement made following a given intervention (Mackey & Gass, 2005; Marsden & Torgerson, 2012; Norris & Ortega, 2000). With regard to the present study, the Control group received no exposure to the target feature either before the study commenced or between pre- and post-test. This level of control was possible due to the researcher teaching German to the participant classes for the full academic year in which the study

[74] The TE-FMC and TE-F groups' delayed post-test scores remained significantly higher than at pre-test

took place (section 3.3). The only exception was the exposure offered within the tests themselves, indicating that the small amount of change observed in the Control group could potentially be attributed to learning as a result of completing the tasks multiple times (Marsden & Torgerson, 2012). Nevertheless several findings are of note. First, no significant change was observed in the Control group's performance between pre- and post-test, with the exception of the (written) Gap-fill and (aural) Act-Out Comprehension tasks. Considering the Gap-fill task first, this significant improvement reflected a very small change in the Control group's score between pre- (M = 11.16, SD = 2.60) and post-test (M = 11.62, SD = 2.63). Analysis of individual test items revealed no evidence of the Control group correctly producing the target feature, ***den*** (object article), within either SVO or OVS constructions. Rather, the overall increase in the Control group's scores at post-test was due to slight improvement in their accuracy in producing the subject article ***der*** under the two test conditions (SVO, OVS). Similarly on the Act-Out Comprehension task the overall significant increase in the Control group's score can be accounted for by a small, yet significant, improvement in comprehension of the SVO test items[75]. There was no improvement in post-test performance on the target OVS sentences.

Based on these findings, the small effect size associated with the Control group did not reflect any improvement in this group's comprehension of the target feature within reversed word order sentences or in their production of the target feature. These findings suggest that no learning (of the target feature ***den***) had occurred in the Control group during the testing phases which, in turn, substantially reduces the likelihood of the TE-FMC and TE-F groups' improvement being due to a test-effect (Campbell & Stanley, 1966; Marsden & Torgerson, 2012).

It is important to acknowledge, however, that whilst exposure to the tests alone did not appear to induce learning in the Control group, the post- and delayed post-tests would have served as extra input for the TE-FMC and TE-F learners following the explicit instruction they received through their respective interventions. Consequently this extra exposure may have contributed to their improved performance. Nevertheless, the observed equivalence of the three groups at pre-test, coupled with the lack of improvement in the Control group on the target items at post-test, suggest that it was the respective interventions which accounted for a majority of the learning gains made by the TE-FMC and TE-F groups at post- and delayed post-test.

[75] SVO test items: pre-test M = 8.35, SD = 1.06; post-test M = 8.78, SD = 0.42

6.2 RQ1: Does explicit grammar instruction improve young learners' a) comprehension and b) production of the target grammatical feature?

A contested issue in effect of instruction research is whether the provision of *input-based* instruction results only in an improvement in learners' ability to comprehend the target feature (e.g. DeKeyser et al., 2002; DeKeyser & Sokalski, 1996), or whether it can also impact their ability to produce the grammatical feature in question (e.g. Benati, 2001; J. F. Lee & Benati, 2007; Stafford et al., 2012; VanPatten, 2002; VanPatten & Cadierno, 1993; Wong, 2004a). In order to contribute to this debate the present study utilised measures of both comprehension and production ability, the findings of which are discussed in the following sections.

6.2.1 Effect of instruction on comprehension

6.2.1.1 Performance on comprehension tasks

A similar pattern of results were observed across the two comprehension tasks (Sentence Matching, Act-Out Comprehension); both the TE-FMC and TE-F groups made significant improvement in their performance at post-test, which was sustained over the delayed post-test. These findings are in line with previous studies which have demonstrated gains in comprehension ability following input-based instruction such as PI (e.g. Benati, 2001; Culman et al., 2009; DeKeyser & Sokalski, 1996; Marsden, 2006; VanPatten & Cadierno, 1993; VanPatten & Wong, 2004).

One finding worthy of note is that the two interventions resulted in equivalent gains in comprehension ability. This is contrary to previous research which has found superior gains in comprehension for PI when compared to alternative forms of input- (Marsden, 2006; Marsden & Chen, 2011) or output-based (e.g. Benati, 2001, 2004; Benati, 2005; Cheng, 2004; VanPatten & Cadierno, 1993; VanPatten & Wong, 2004) instruction (see section 6.4 for further discussion).

6.2.1.2 Overreliance on word order

With regard to learner performance by test item (SVO, OVS), on both comprehension tasks all three groups were performing at ceiling level on the SVO condition from pre-test, but were incorrectly interpreting OVS structures. It can therefore be argued that at pre-test the learners were relying primarily on word order and interpreting the first noun as the subject and the second noun as the object. This is in line with previous research which has demonstrated an overreliance on word order by L1 English learners when interpreting L2 German input (Culman et al., 2009; Jackson, 2007; LoCoco, 1987; VanPatten & Borst, 2012). It can be argued that this overreliance was due to the learners encountering a processing problem, such as that predicted by VanPatten's (1996, 2004a) FNP (section 2.3.3.2), which impeded their interpretation of the target grammatical feature (***den***). Alternatively, it could be argued, in line with the Competition Model (MacWhinney, 1987, 2001, 2005), that the learners were transferring the cue most reliable in their L1 English, i.e. word order, to their interpretation of target language sentences.

Analysis of the learners' performance across the Animacy conditions (A+A, A+I, I+A) on the Act-Out Comprehension task provided further evidence of their reliance on word order at pre-test. It was hypothesised that if the learners were attending to animacy as well as word order, then: A) there would be evidence of their correctly interpreting OVS sentences which contained an Animate subject and Inanimate object (A+I) at pre-test; and B) across both word orders (SVO, OVS) the largest gains would be made on sentences containing an Inanimate subject and Animate object (I+A). Contrary to these hypotheses, analysis revealed minimal differences between the three Animacy conditions both in terms of overall scores (Table 4.16) and when broken down by test condition (SVO, OVS) (Table 4.17, 4.18). Furthermore, there were no differences in the gains made on the respective Animacy conditions between pre- and post-test (Table 4.19). These findings suggested that the learners were paying little attention to animacy cues when interpreting target language sentences. In addition, on the OVS items, both the TE-FMC and TE-F groups made substantial, equivalent gains across all of the Animacy conditions (Table 4.19), reflecting the improvement made in the learners' comprehension of the target feature and reversed word order sentences.

Previous research has demonstrated that animacy cues may also be utilised by L1 English speakers (e.g. Kempe & MacWhinney, 1998) and, in some cases, can overrule word order cues (e.g. Jackson, 2007) when the two are in conflict. In their picture choice task, Kempe and MacWhinney (1998) utilised an online measure of decision speed and observed that learners responded more quickly when the first noun in a sentence was animate or when the second noun was inanimate. Furthermore, Jackson (2007) found that from pre-test participants were consistently correctly interpreting sentences comprising one animate and one inanimate noun, regardless of whether the sentences were presented in SO or OS order. Similarly in his *Lexical Semantics principle* VanPatten (2004a, 2007) acknowledged the potential influence animacy can have on the processing of target language input, arguing that factors such as animacy may attenuate the FNP in certain circumstances. Nonetheless, this proposal was not borne out in the present study; the learners universally overlooked animacy cues, relying solely on word order to aid interpretation at pre-test. Notably, in Jackson's (2007) study it was the animacy of the subject noun that was manipulated, rather than that of the object as in Kempe and MacWhinney's (1998) study. Jackson (2007) therefore proposed that the observed effects of animacy may stem from "the ease with which … one can interpret the target sentence based on the semantic meaning of each individual word in the sentence, as well as how the words fit together according to real-world knowledge" (p. 425). This potential influence of "real-world knowledge" is also captured in VanPatten's (2004a, 2007) *Event Probability principle* (section 2.3.3.3). With regard to the present study, it can be argued that the potential influence of event probabilities and real world knowledge may have been lessened through the use of cartoon pictures (Sentence Matching task) and toys (Act-Out Comprehension task), which allowed learners to interpret the sentence stimuli based on the premise that 'anything goes'. In turn, this may have mitigated the potential impact of animacy on learners' comprehension of the target language input, thereby explaining the lack of animacy effects observed.

6.2.1.3 Overcoming an overreliance on word order

By post-test both the TE-FMC and TE-F groups had made significant improvement in their interpretation of the OVS structures on both comprehension tasks (Tables 4.2, 4.6). Given the learners' ceiling level performance on the SVO items, this significant improvement in their scores on the OVS items can account for the overall improvement in the TE-FMC and TE-F groups' performance on these tasks. Further, these results provide evidence that, following their respective interventions, both the TE-FMC and TE-F groups had significantly improved in their interpretation of the target feature when it occurred in sentences with reversed word order. These findings are in line with previous research (Culman et al., 2009; Jackson, 2007; VanPatten & Borst, 2012) which has demonstrated that L1 English learners can be successfully trained to attend to case-marking cues when interpreting L2 German input. The present study has now extended this finding to young learners. Based on Input Processing theory (VanPatten, 2004a, 2007) this improvement in the TE-FMC and TE-F learners' comprehension of OVS structures could be taken as evidence that instruction successfully aided the learners in overcoming the processing problem denoted by the FNP, namely a strict reliance on SVO word order. Successfully pushing the learners to abandon this less-than-optimal strategy enabled them to correctly connect the target feature (***den***) with its meaning within the input (i.e. marking the object). This could, as argued by VanPatten (2004a, 2007) and Wong (2004a), have created richer intake for the developing system which, in turn, resulted in the subsequent acquisition of the target feature, as evidenced by the learners' improved comprehension ability.

Nevertheless, alternative explanations can also be put forth. Under the Competition Model, learning to attend to a novel grammatical cue (i.e. one which is not part of the L1) entails a restructuring of the learner's cue hierarchy (MacWhinney, 2001). Therefore the TE-FMC and TE-F groups' performance on the comprehension tasks could be taken as evidence that such a restructuring had occurred. At post-test and delayed post-test, the learners demonstrated reliance on case-marking when assigning grammatical roles to target sentence stimuli, which had superseded their previously dominant L1 word order cue. Additionally on the oral Act-Out Production task, at both post- and delayed post-test, there were several instances of a sub-sample of the TE-FMC and TE-F learners utilising OVS word order in the sentences they produced (Appendix 11). Stafford et al. (2012), investigating the learning of case and agreement cues in L2 Latin, observed a similar reduction in the frequency of SVO word order being marked by their L1 English learners in a post-test production task. Given the importance of word order for the assignment of grammatical roles in English, such a finding could provide further evidence that a restructuring of the learners' cue hierarchy may have occurred (Stafford et al., 2012).

6.2.2 Effect of instruction on production

6.2.2.1 Performance on production tasks

As with the comprehension activities, a consistent pattern of results was observed across the written (Gap-fill) and oral (Act-Out Production, Sentence Repetition) production tasks: significant improvement in both the TE-FMC and TE-F learners' performance at post-test[76], which was maintained at delayed post-test; no difference between the TE-FMC and TE-F groups at post- and delayed post-test; and no improvement in the Control group's scores at post-test. Notably, however, on the Sentence Repetition task the difference between the TE-FMC and TE-F groups approached significance at delayed post-test, with the TE-FMC group outperforming the TE-F group (discussed in section 6.4.3).

Overall findings for the production measures are consistent with previous research (e.g. Agiasophiti, 2013; Benati, 2001; Marsden, 2006; Marsden & Chen, 2011; Stafford et al., 2012; VanPatten & Cadierno, 1993; Wong, 2004b) which has demonstrated that input-based instruction can improve learners' ability to produce the target grammatical feature, even though the instruction itself provides no overt production practice. Such studies have

[76] Gap-fill: TE-FMC d = 1.73, TE-F d = 1.79; Act-Out Production: TE-FMC d = 1.62, TE-F d = 1.24; Sentence Repetition: TE-FMC d = 1.45, TE-F d = 1.27

argued that input-based instruction (primarily PI) enhanced the way in which learners processed target language input and, consequently provided intake for the developing system which, in turn, resulted not only in better “processing of input” but also “knowledge that is apparently also available for production” (VanPatten & Cadierno, 1993, p. 240). Indeed, Wong (2004a) proposed that an improvement in learners’ productive ability, following instruction which does not involve any output practice, provides evidence that “some kind of change” (p. 52) in their interlanguage system has occurred, which is accessible during production as well as comprehension. Several studies (Farley, 2004a; Rosa & Leow, 2004b; Stafford et al., 2012; Wong, 2004b) have provided evidence that the explicit information received during the intervention may have contributed to learners’ improved performance on the production measures. Stafford et al. (2012) observed that, whilst all of their treatment groups (varied by degree of explicitness) made gains in their production of the target feature[77], the greatest gains were made by the group who had received both explicit pre-practice grammar explanation and explicit feedback. Similarly, Rosa and Leow (2004b) found that those learners who had received a more explicit form of instruction (+ Explicit information and/or + Explicit feedback) made the greatest improvement in their production of Spanish conditional sentences, although the learners in the more implicit conditions also exhibited improvement. Based on such findings it could be argued, then, that the learners’ improved productive ability may be due to the development of functional explicit knowledge of the target feature. Indeed Rosa and Leow (2004b) proposed that “those conditions that favoured the development of explicit knowledge (i.e. those containing some form of explicit linguistic information) were more effective than less explicit conditions in helping learners to improve their ability to produce items” (p. 208).

In the present study, both the TE-FMC and TE-F groups received the same explicit information relating to the target feature, which was repeated on three occasions throughout the intervention. In addition the TE-FMC group received some explicit feedback relating to the target FMC during the activities. In line with the above findings it could be argued that the TE-FMC and TE-F groups’ improved productive ability was due to the explicit information they had received and the subsequent development of explicit knowledge. However, since the present study did not include a treatment group who completed the intervention activities (TE-FMC or TE-F) but did not receive explicit information, it is not possible to draw such a conclusion (see sections 6.3 and 6.4.2.2 for further discussion).

[77] Assignment of agent/patient roles, based on verb agreement, word order, and case (nominative/accusative) in Latin

Notably, a number of studies have produced contrary evidence regarding the respective benefits of input- (e.g. PI) versus output-based (e.g. TI) practice (e.g. Allen, 2000; DeKeyser & Sokalski, 1996, 2001). DeKeyser and Sokalski (1996, 2001) argued that the findings of their 1996 self-labelled replication of VanPatten and Cadierno's (1993) study reflected the predictions of Skill Acquisition Theory (SAT); namely that input practice is of most benefit for comprehension skills and that output practice most benefits production skills. In line with SAT, DeKeyser and Sokalski (1996, 2001) argued that any observed gains in production following input-based instruction (either in their study or in studies such as VanPatten & Cadierno, 1993) were likely due to a transfer of declarative knowledge rather than evidence of proceduralization or automatization having occurred. Additionally, in their comparison of PI and traditional output-based instruction (TI), Allen (2000) observed equivalent learning gains for the two groups on a comprehension task, however an advantage for TI over PI in terms of production. Arguably the present study's findings could be interpreted as contrary to these studies since the TE-FMC and TE-F interventions, both of which are input-based only, resulted in the learners making substantial gains in their ability to produce, as well as comprehend, the target feature as measured on both written and oral tasks. Nevertheless it is important to acknowledge two key points. First, whilst there was a significant increase in their scores at post-test, the TE-FMC and TE-F learners did not universally reach ceiling level on the three production tasks. Secondly, the present study did not draw any comparison between the TE-FMC and TE-F interventions and an alternative, comparable form of output-based instruction. Therefore, in line with the predictions of SAT, it is possible that an output-based intervention may have led to even greater gains on the production measures. Nonetheless the benefits of the input-based instruction utilised in the present study (TE-FMC and TE-F) for both comprehension and production cannot be denied.

6.2.2.2 Use of the target feature *den*

Analysis by test item illuminated that the overall improvement exhibited by the TE-FMC and TE-F learners could be attributed to the significant, sustained improvement in their use of the target feature ***den***; for the Gap-fill and Sentence Repetition tasks this improvement was observed on the test items in both SVO and OVS word orders (Tables 4.4, 4.8, 4.12). In contrast, the Control group's performance on the test conditions targeting production of the object article ***den*** remained at baseline.

Remarkably, to date, those research studies which have been concerned with nominative-accusative case marking in German have focused almost exclusively on learners' *comprehension* of this grammatical feature. The respective studies carried out by Jackson

(2007) and LoCoco (1987), for instance, examined whether and how L1 English learners of L2 German made use of case-marking cues (along with word order and animacy cues) when interpreting target language input. Kempe and MacWhinney (1998) utilized a computerized picture choice task to compare the comprehension of case-marking cues by L1 English learners of L2 German and Russian. Furthermore, a series of studies (Culman et al., 2009; Henry et al., 2009; VanPatten & Borst, 2012) have utilised an online measure (trials to criterion) to determine how soon L1 English learners of L2 German began processing case-marking correctly in referential PI activities. These studies were primarily concerned with isolating the effectiveness of two components of the PI package: explicit information and referential activities.

Only one study (Agiasophiti, 2013) has been identified as measuring learners' production (via a written fill-the-gap task), as well as comprehension, of the target feature. In a comparison of the effects of enhanced versus unenhanced PI[78], Agiasophiti (2013) observed significant gains across all treatment conditions at post-test. This finding is in line with that of the present study, although the present study utilised oral as well as written production measures. Contrary to the present study, however, Agiasophiti (2013) observed a substantial decrease in learners' performance on the fill-the-gap task at delayed post-test, although it is not reported whether this was significant. Notably Agiasophiti (2013) utilised a short intervention (as defined by Norris & Ortega, 2000, p. 477) of only 1.5 hours over two days, whereas the treatment in the present study was considerably longer (4 hours 10 minutes over five weeks). The prolonged exposure to and practice of the target feature afforded during the five-week intervention in the present study may, therefore, provide an explanation as to why there was a discrepancy in delayed post-test findings between the two studies.

Overall, the present study has provided robust evidence that the TE-FMC and TE-F learners were able accurately to produce the target feature ***den*** when tested on both written and oral production tasks.

6.2.2.3 Overuse of the target feature *den*

During the process of acquiring a new grammatical feature the learner may go through a period of overgeneralisation in which they overuse the new feature in non-obligatory contexts (Long, 2010; Long & Robinson, 1998). This could be interpreted as a reflection of underlying "hypothesis-testing" which occurs as part of the learning process (Long, 2010, p.

[78] All components of the PI package were included in both treatments; explicit information, referential and affective activities. Enhancement (of all components) involved colour coding of nominative / accusative case-marked articles (Agiasophiti, 2013, p. 167)

379). Indeed, an increase in errors may constitute an inevitable prerequisite to the assimilation of a new feature, resulting in a U-shaped trajectory of learning (Kellerman, 1985; Long, 2010; Long & Robinson, 1998). With regard to the present study, despite significant improvement in the TE-FMC and TE-F learners' use of the target feature ***den***, it was important to explore whether any reduction had occurred in the learners' use of the subject article ***der*** at post- and delayed post-test, reflecting an overuse of the newly learned article (Truscott, 1999, 2005).

A number of instances of learners overgeneralising their use of ***den*** were observed across the three production tasks at post-test. First, on the written Gap-fill task a small, significant decrease was found in the TE-FMC groups' scores on both the SVO+Subj and OVS+Subj conditions at post-test, due to a reduction in the learners' production of ***der*** (replaced by ***den***). Secondly, on the oral Sentence Repetition task a significant increase (Tables 4.14, 4.15) was observed at post-test in both the TE-FMC and TE-F groups' overuse of ***den*** on the test items targeting the subject article (SVO+Subj, OVS+Subj). A similar finding was observed for the Act-Out Production task in which a small, albeit significant, increase in both groups' overuse of ***den*** was again observed (Table 4.10, 4.11). Notably, the increase in errors (due to overuse of ***den***) made by the TE-FMC learners on the subject conditions of the Sentence Repetition and Act-Out Production tasks had decreased by delayed post-test. The TE-FMC group's performance across the three time points was, therefore, in line with the proposed U-shaped trajectory of learning (Kellerman, 1985; Long, 2010; Long & Robinson, 1998). In contrast, for the TE-F group, an increase in errors was observed again at delayed post-test for both the Sentence Repetition and Act-Out Production tasks. This suggests that the TE-F learners may not yet have moved beyond the 'dip' in performance characterised by an increase in errors and overuse of the target feature. It is important to acknowledge that no significant difference was found between the two groups across any of the tasks or conditions. Nonetheless the above findings suggest a tendency within the TE-F group to continue to overgeneralise their use of the new article ***den*** at delayed post-test, a tendency which had significantly reduced in the TE-FMC group by the delayed post-test. This is discussed in more detail in section 6.4.3.

Notably, the largest amount of overgeneralisation was observed on the Sentence Repetition task. For the SVO+Subj condition the TE-FMC group overused ***den*** in 18.5% of cases and the TE-F group in 13%, whereas on the Act-Out Production task the percentage use of ***den*** for the subject condition was 5.7% and 4.5% for the two groups respectively. This is likely to be a reflection of the nature of the task itself since, in the Sentence Repetition task which was designed in line with the general guidelines for an Elicited Imitation task

(Erlam, 2006), the learners were required to repeat sentence stimuli following a short beep. Use of both SVO and OVS sentence structures would have exposed the learners to the two articles (***der*** and ***den***) in both sentence initial and medial positions which, in turn, may have served to prime some learners to a certain extent when they came to repeat the sentences. Indeed, the incidence of the Control group incorrectly producing the object article ***den*** in the subject condition was also higher on the Sentence Repetition task than on the Act-Out Production task (Table 4.10, 4.14). Notably, this overuse of ***den***, coupled with the lack of improvement in the Control groups' production of ***den*** compared to that of the TE-FMC and TE-F groups, suggests that the learners were not simply relying on rote repetition when completing this task.

6.2.3 Written versus oral tasks

One notable finding of the present study was that equivalent learning gains were found across both the written and oral tests of comprehension and production. This demonstrates that not only can explicit input-based instruction result in improvements on written, sentence-level tasks, as has been shown in several previous PI studies (e.g. Benati, 2001, 2004; Benati, 2005; Cadierno, 1995; Farley, 2004a, 2004b; VanPatten & Cadierno, 1993; VanPatten & Oikkenon, 1996; VanPatten & Wong, 2004), but also that it can result in sustained improvement in learners' use of the target grammatical feature on measures of more spontaneous communication ability. This constitutes a key finding given that a majority of PI studies, such as those cited above, have primarily employed written rather that oral outcome measures.

The oral comprehension measure utilised in the present study constituted a test of listening ability. Notably this test was different to the listening comprehension tests employed in previous PI studies, in that the learners were required to respond immediately by acting out the aural stimuli using soft toys rather than, for example, selecting from two or more options on a written worksheet (e.g. Agiasophiti, 2013; Benati, 2001; Marsden, 2006; VanPatten & Cadierno, 1993). This constituted a more time and communicatively pressured task. Despite this operational difference, substantial durable learning gains were still observed for both the TE-FMC and TE-F learners, which is in line with the consistently positive gains observed on comprehension tests employing a written response.

In terms of the learners' ability to produce the target feature, a number of previous studies have utilised discourse-level tasks in a bid to provide a more robust measure of learners' productive ability; however these have tended to constitute written, rather than oral, composition tasks (e.g. Cheng, 2004; Sanz, 2004; Sanz & Morgan-Short, 2004). Studies that have utilised an oral measure of production have yielded mixed results. Marsden (2006), for example, observed significant improvement on an oral production task for one class following instruction (PI), whereas other studies have found no, or minimal, effect of PI on learners' performance on oral production measures (e.g. Marsden, 2006 (class 2, compared to the Enriched Input group); Marsden & Chen, 2011; VanPatten & Sanz, 1995). In the present study, however, a significant and sustained improvement was revealed for both the TE-FMC and TE-F learners' on the two oral measures of production.

The use of tasks in both written and oral modalities goes some way to addressing the concern levelled at some PI studies, as well as effect of instruction research more generally, that relatively few measures of more spontaneous communication ability (e.g. discourse-level tasks, oral tasks, tasks including increased time pressure) have been utilised (Doughty, 2003, 2004; Norris & Ortega, 2000). It is important to acknowledge that the oral tasks employed in the present study were sentence- rather than discourse-level, since the use of discourse-level tasks was not appropriate given the L2 ability (beginner) of the learners. Nevertheless, by virtue of the fact that the activities were completed one-to-one with the researcher and utilised soft toys and pictures to keep the learners focussed on meaning, the oral tasks were intended to place a greater time and communicative pressure on the learners than was experienced in the written activities (R. Ellis, 2002, 2005, 2009b). Indeed, as proposed by Trahey and White (1993), it was anticipated that "under pressure to perform orally, students would have less time to think about their responses than in the written … tasks" (p. 190) (see section 6.3.3.2 for further discussion).

Finally, it was deemed necessary to employ both written and oral comprehension and production measures to maintain the ecological validity of the present study. As a classroom-based study it was crucial to determine the effectiveness of the TE-FMC and TE-F interventions for developing the learners' ability to employ the target feature in all four skills, reading (written, comprehension), writing (written, production), listening (aural, comprehension) and speaking (oral, production); as such both interventions appear to have been successful.

6.2.4 Summary of discussion for RQ1

In response to RQ1, the answer is 'yes'; explicit, input-based grammar instruction (i.e. TE-FMC and TE-F interventions) can result in substantial, durable improvements in learners' comprehension and production of definite article case-marking in German. Based on these findings it is possible to conclude that both interventions were successful in pushing the learners to attend to the target grammatical feature and, crucially, its FMC, resulting in the target feature being assimilated into the learners' interlanguage, as evidenced by the learners' post- and delayed post-test performances. Further, both interventions proved successful in assisting the learners' in overcoming their reliance on word order which, as predicted by both the FNP (VanPatten, 1996, 2004a) and Competition Model (MacWhinney, 2001; MacWhinney & Bates, 1989), had dominated the learners' interpretation of target language input at pre-test and continued to do so for the Control group at post-test. In addition, the TE-FMC and TE-F learners exhibited significant improvement both on the written tasks, in which the learners had time to reflect on their language use, and the oral tasks. This finding suggested that the target feature had been integrated into the learners' interlanguage in a way which allowed them to perform equally well under much greater time- and communicative pressure.

6.3 RQ2: To what extent does explicit grammar instruction develop different types of knowledge of the target grammatical feature?

The PCA of the TE-FMC and TE-F groups' scores on the six outcome measures suggested that at post-test the written and metalinguistic tasks were eliciting a different type of knowledge to that of the oral tasks. This was in line with the way in which each task was designed; namely that the written and metalinguistic tasks were designed to elicit more explicit knowledge, whereas the oral tasks were designed to elicit more implicit knowledge. At this point it is useful to remind the reader of the way in which 'explicit' and 'implicit' knowledge have been defined and operationalised for the purposes of the present study, i.e. based on three of the criteria put forward by R. Ellis (2009b); *time available*, *focus of attention* and *meta-language* (see section 2.2.6).

Contrary to the post-test findings, at delayed post-test all of the tasks were found to load as one component. Coupled with the finding that the learners' performance on the Sentence Reconstruction task (test of metalinguistic knowledge) had significantly decreased at delayed post-test, this suggested that all six tasks were tapping into a similar or related type of knowledge at this time point. Within the context of these findings the following sections will discuss the evidence available to inform the question of whether, following their respective interventions, the TE-FMC and TE-F learners had not only developed explicit knowledge, but also more implicit knowledge, of the target feature.

6.3.1 Evidence of explicit knowledge

The results of the present study support the proposal that input-based instruction, consisting of explicit information followed by activities which involve some kind of task-essential attention to form, will result in explicit learning and the development of explicit knowledge of the target grammatical feature (De Jong, 2005; DeKeyser et al., 2002; Doughty, 2004; Marsden & Chen, 2011). A number of findings provide support for this claim. Firstly, consistently large correlations (.46 < *r* < .75) were observed between the two written measures across the post- and delayed post-test. Secondly, as detailed above, at post-test the learners' performance on the written and metalinguistic tasks was found to pattern separately to the oral measures on the PCA. These findings suggest that, at post-test, these tasks were tapping into a similar type of knowledge which, by virtue of their design, was likely to be more explicit. Furthermore, the written Sentence Matching and Gap-fill tasks were untimed, sentence-level tasks, which would have allowed the learners the opportunity to monitor and reflect on their comprehension and use of the target feature (De Jong, 2005; Doughty, 2003, 2004; R. Ellis, 2005, 2009b). Such tasks lie at the 'decontextualised' end of the continuum (from decontextualized to naturalistic) along which Chaudron (2003) has classified the data collection tools commonly utilised in SLA research. As such the written tasks were arguably more likely to draw on explicit, declarative knowledge, since within such a context language use is highly constrained and controlled (Doughty, 2003, 2004; R. Ellis, 2005, 2009b; Norris & Ortega, 2000).

It is important to note, however, that most tasks are likely to draw on both explicit and implicit knowledge to a certain extent (DeKeyser, 2003; R. Ellis, 2002, 2005; Roehr, 2008); therefore the possibility of the learners' utilising more implicit knowledge in the written tasks cannot be discounted (as noted in Marsden & Chen, 2011). Nevertheless, the finding that the written tasks patterned together with the metalinguistic task on the PCA at post-test supports the proposal that the learners were relying on more explicit knowledge when completing these tasks at post-test. The next section discusses the evidence provided by the Sentence Reconstruction task that this knowledge was not only explicit but also verbalisable.

6.3.2 Evidence of metalinguistic knowledge

i) At post-test

Quantitative (section 4.6.1) and qualitative (section 4.6.2) analyses of the TE-FMC and TE-F learners' explanations in the Sentence Reconstruction task demonstrated that, by post-test, the learners had developed metalinguistic (i.e. explicit, verbalisable) knowledge of the target grammatical feature. A majority of learners could explain their use of the target feature, in many cases utilising appropriate metalinguistic terminology (section 4.6.2.2). Even those learners who were unable to draw on metalinguistic terminology succeeded in explaining the relevant grammatical rule in their own words. These findings are in line with previous studies (e.g. Barton et al., 2009; Bouffard & Sarkar, 2008; Roehr, 2012) which have demonstrated that young learners are able to express awareness of language and engage in linguistic analysis. In contrast, no such evidence was ascertained from the Control group learners at pre- or post-test, illustrating that they were neither aware of nor could articulate the target grammatical rule.

Many of the TE-FMC and TE-F learners also demonstrated that they were able to articulate their understanding of the differential case-marking of masculine versus feminine and neuter articles (i.e. only masculine articles are marked in the accusative case). These learners demonstrated that they were able to apply their knowledge of ***den*** in order to correctly interpret sentences containing one non-case-marked (feminine or neuter) noun. This was further illustrated by the learners' improved performance on the two written tasks, in which all of the sentence stimuli contained one masculine plus one feminine or neuter noun. Additionally, in the Sentence Reconstruction task some learners were able to explain how they utilised their knowledge of the target feature in order to correctly interpret sentences in reversed word order. These observations provide further evidence that a "restructuring of [the learners'] cue hierarchies" (Stafford et al., 2012, p. 764) had occurred, which allowed the learners to circumvent their L1 word order rule and rely instead on the more reliable (in German) case-marking cue when interpreting sentence stimuli.

Nevertheless there was a sub-group of both the TE-FMC (N = 4 of 45) and TE-F (N = 8 of 41) learners who were unable to articulate the target grammatical rule at post-test. These learners relied instead on explanations related to gender, animacy and guesswork, as all of the learners had at pre-test and as the Control group continued to do at post-test. R. Ellis (2009a) stated that differences in learners' analytical skills (such as those needed to memorise, induce or deduce explicit knowledge) can constrain an individual's ability to learn explicit facts about a language. Such differences could, therefore, account for why a small number of learners were unable to provide correct explanations relating to the target feature on this task. In general, a lack of metalinguistic knowledge, as evidenced by low (t = 0 to 2) performance on the Sentence Reconstruction task (i.e. Not Got It sub-group), tended to be associated with a lack of improvement across the other five outcome measures. For example, both learners in Table 6.1 (over page) were performing at chance level on all of the outcome measures at post-test:

Table 6.1: Post-test scores for two learners from the Not Got It sub-group of Sentence Reconstruction (Explanation) task

Participant	Group	S. Recon (E) (*k* = 6)	S. Match (*k* = 24)	Gap-fill (*k* = 24)	AO Comp (*k* = 18)	AO Prod (*k* = 24)	S. Rep (*k* = 12)
65	**TE-FMC**	**2**	10	8	10	11	8
85	**TE-F**	**1**	10	11	9	12	9

Notably, however, there were a number of exceptions to this finding. A handful of TE-FMC (*N* = 3) and TE-F (*N* = 3) learners who were categorised as Not Got It on the Sentence Reconstruction (Explanation) task were found to have scored at or close to ceiling level on some, although not all, of the written and / or oral tasks at post-test (see Appendix 25). In their study Green and Hecht (1992) observed a similar discrepancy between learners' ability to talk about versus use a given grammatical rule with a set of L2 English sentences. For 70% of instances in which Green and Hecht's participants were unable to supply the correct grammatical rule, the participants (L1 German) were, nonetheless, able to correct the grammatical errors.

For the Sentence Reconstruction task, the Not Got It sub-group constituted only a small percentage of learners within each group (TE-FMC, 9%; TE-F, 20%), suggesting that a majority of the learners were able to demonstrate metalinguistic knowledge of the target grammatical feature. In contrast, the percentage of learners within the Not Got It sub-group was larger across the written and oral comprehension and production tasks than for the Sentence Reconstruction task. This indicated that the ability to explain the grammatical rule in question did not necessarily equate with the ability to comprehend and use the target feature correctly in the other tasks. For example, both learners in Table 6.2 performed at ceiling level on the Sentence Reconstruction task at post-test, yet continued to score at chance (i.e. baseline) level on the remaining five tasks.

Table 6.2: Post-test scores on the six outcome measures for two learners from Got It sub-group of Sentence Reconstruction (Explanation) task

Participant	Group	S. Recon (E) (k = 6)	S. Match (k = 24)	Gap-fill (k = 24)	AO Comp (k = 18)	AO Prod (k = 24)	S. Rep (k = 12)
15	**TE-FMC**	**6**	14	12	7	12	6
35	**TE-F**	**6**	12	10	9	12	7

Harley (1998) observed a similar disparity between talking about versus using the target feature and noted that "some children were consciously aware of the relevance of noun endings for gender attribution even if they had not spontaneously produced correct articles during the testing" (p. 168). These findings are in line with the proposal that "knowing a language rule does not mean one will be able to use it in communicative interaction" (Lightbown, 2000).

Despite the fact that ability to talk about the target grammatical rule did not *guarantee* improved comprehension and production of the target feature, those learners who did show improvement across the outcome measures at post-test were also consistently found to be performing at ceiling level on the Sentence Reconstruction task. Therefore explicit knowledge of the grammatical feature, gained through the explicit information provided within the intervention, was generally associated with an improvement in comprehension and production. This finding is in line with previous studies, which have demonstrated greater levels of learning for those learners who were 'aware' of (i.e. able to verbalise) the target grammatical rule than for those who were 'unaware' (i.e. unable to verbalise the rule) (e.g. Hama & Leow, 2010; Robinson, 1997; Rosa & Leow, 2004a). Indeed, even in studies which have claimed to provide evidence of learning without awareness, greater learning gains have still tended to be observed for learners who have demonstrated awareness (e.g. Leung & Williams, 2012; Williams, 2005).

It is important to note that the association between evidence of metalinguistic knowledge and improvement across the written and oral tasks was likely to have been mediated to a certain extent by individual differences. Pienemann (1989), for example, has argued that, whilst instruction can serve to speed up acquisition, this effect will only be evident if a learner is developmentally ready to acquire the grammatical feature in question. Similarly, previous research has demonstrated that factors such as aptitude (Harley & Hart, 2002; Ranta, 2002) and more general language proficiency might also influence the effectiveness of instruction for individual learners. Such differences could, therefore, account for why some learners (Table 6.2) were able to articulate the grammatical rule presented during the interventions, however did not demonstrate any evidence of having improved in their comprehension or production of the target feature (see section 6.4.5 for further discussion).

ii) At delayed post-test

Whilst analysis revealed that the TE-FMC and TE-F groups' performance on the Sentence Reconstruction (Explanation) task was significantly higher at delayed post-test than at pre-test, a significant decrease in learners' scores occurred between post- and delayed post-test. Examination of the learners' explanations (section 4.6.2.3) illuminated the nature of this decrease. First, many learners (TE-FMC, N = 15; TE-F, N = 17) were unable to articulate the target rule at delayed post-test, despite having been able to do so post-test. Secondly, there were inconsistencies between the explanations given for individual test items; at delayed post-test a number of learners (TE-FMC, *n* = 11; TE-F, *n* = 16) gave the correct explanation on one test item, however were unable to explain their use of the target feature in the next test item. This inconsistency was reflected in the observation that some learners could explain the function of the target feature within test items containing two masculine nouns, however struggled in extrapolating this knowledge to those test items containing one masculine plus one feminine or neuter noun.

This decline in the learners' ability to talk about and explain the function of the target grammatical feature was also reflected in the analysis of individual learners' performance on this task. Notably, for the metalinguistic task, the percentage of learners within the Not Got It sub-group increased between post- and delayed post-test[79], whereas the percentage within the Got It group decreased[80]. This change in the distribution of learners across the sub-groups reflected the significant decrease in the learners' scores between post- and delayed post-test.

These observations illustrate the fragile nature of the learners' metalinguistic knowledge at delayed post-test and the difficulty the learners had in articulating that knowledge. The findings are in line with the suggestion that explicit knowledge is susceptible to decay over time (Mitchell, Myles, & Marsden, 2012). Nevertheless, it is important to note that a decrease in scores on the Sentence Reconstruction task does not necessarily reflect a complete loss of explicit knowledge but, rather, a reduction in the learners' ability to articulate their knowledge of the target feature nine weeks after the intervention.

6.3.3 Evidence of a different type of knowledge

6.3.3.1 A more implicit form of knowledge?

Analysis of the written and metalinguistic tasks indicated that the learners had developed explicit knowledge of the target grammatical feature. A more contentious issue, however, is whether the interventions utilised in the present study also led to the learners developing a more implicit form of knowledge. Indeed, the issue of whether it is possible to develop more implicit knowledge following explicit instruction has received considerable attention within the literature (e.g. DeKeyser, 2003; DeKeyser & Criado, 2012; Doughty, 2003; N. Ellis, 2005, 2010; R. Ellis, 2002, 2006; Krashen, 1982; Marsden & Chen, 2011) and, as yet, remains unresolved.

---

[79] Not Got It: TE-FMC group, 9% to 33%; TE-F group, 20% to 41%

[80] Got It: TE-FMC group, 60% to 42%; TE-F group, 60% to 20%

With regard to the present study a number of findings can contribute to this discussion. First, analysis of the learners' test scores via PCA revealed that the oral tasks patterned separately to both the written and metalinguistic tasks at post-test. This was a similar finding to that of Marsden and Chen (2011): their oral tasks loaded separately to a written grammaticality judgement task and gap-fill task when analysed via PCA. Secondly, no correlation was found between the learners' scores on the oral tasks and their scores on the test of metalinguistic knowledge (Sentence Reconstruction task) at post-test. Thirdly, although significant correlations were observed between each of the oral and written tasks, these were not as strong as between the oral tasks themselves (Table 5.3). Taken together, these findings suggest that the oral tasks were tapping into a different type of knowledge than either the written or metalinguistic tasks at post-test; however the qualitative nature of that knowledge is difficult to define.

Based on the design of the oral tasks it could be argued that they may have been eliciting a more implicit form of knowledge, since the learners were required to complete the tasks under time and communicative pressure whilst focussed primarily on meaning (R. Ellis, 2005, 2009b). As such, the oral tasks would not have afforded the learners much opportunity to consciously reflect on or monitor their language use (R. Ellis, 2005; Roehr, 2008; Trahey & White, 1993). Such an interpretation is in line with previous research (e.g. R. Ellis, 2005; Han & Ellis, 1998) which has demonstrated that oral tasks, such as those utilised in the present study, can constitute a measure of learners' more implicit knowledge, provided that the tasks adhere to the given specifications (e.g. time-pressured, meaning-focussed, etc.) (R. Ellis, 2009b, p. 40).

Furthermore, the PCA conducted on the delayed post-test data can also shed light on the type of knowledge the learners had developed following instruction. In contrast to the post-test data, at delayed post-test all of the tasks were found to load as one component, suggesting that at delayed post-test the tasks were tapping into a more homogenous type of knowledge. Taken together with the observed decrease in the learners' performance on the metalinguistic (Sentence Reconstruction) task, this indicates that the learners may have been relying on a more implicit form of knowledge when completing the tasks at delayed post-test. Such an interpretation would be in line with the observation that, firstly, implicit knowledge is more durable than explicit knowledge (R. Ellis & Shintani, 2013; Mitchell et al., 2012) and, secondly, that implicit knowledge is associated with more consistent test performance (R. Ellis, 2009b). Overall, there were inconsistencies in the learners' performance on the metalinguistic test at delayed post-test, whereas their performance on the written and oral tasks remained consistently high. Additionally, the finding that the written tasks patterned together with the oral tasks at delayed post-test is reflective of the fact that, as expressed by R. Ellis (2005), "learners are likely to draw on whatever resources they have at their disposal irrespective of which resources are the ones suited to the task at hand" (p.153). Although the written tasks promoted the use of explicit knowledge, the results of the PCA indicated that the learners' may have been relying more on a different type of knowledge when completing the Sentence Matching and Gap-fill tasks at delayed post-test.

6.3.3.2 Methodological considerations

Despite the above findings in support of the proposal that the learners may have developed a more implicit form of knowledge of the target feature, it is important to note a number of relevant methodological issues. First, although when completing the oral tasks the learners were under pressure to respond to the researcher immediately, the tasks themselves were not timed. Secondly, features associated with fluent language use, such as pauses and reformulations, were not monitored (R. Ellis, 2002), although, based on the researcher's anecdotal observations, reformulations were infrequent. It is possible, therefore, that some learners may have been monitoring their own performance at times during the tasks, despite the fact that the tasks were designed in order to reduce such behaviour; monitoring by definition would imply the use of some level of explicit knowledge (Krashen, 1982). Indeed, whilst characteristics such as time pressure might *encourage* the use of more implicit knowledge, they do not guarantee the use of one or the other type of knowledge (DeKeyser, 2003). Further, as acknowledged previously, most tasks are likely to draw on both knowledge types to a certain extent (R. Ellis, 2002, 2005; Roehr, 2008).

Additionally, it has been argued that 'free' discourse-level production tasks, i.e. communicative tasks which provide few (or no) constraints on the learners' language use, provide a more robust measure of learners' spontaneous communicative ability and, therefore, implicit knowledge (Doughty, 2003, 2004). With regard to the present study, however, whilst the oral tasks were designed to provide a more meaning-focussed, communicative activity through the use of soft toys and one-to-one interaction with the researcher, they remained characteristic of sentence-level, "constrained constructed response" tasks (Norris & Ortega, 2000, p. 440). This was deemed necessary given the limited vocabulary of the learners and the difficulty of 'trapping' OVS sentences in more spontaneous discourse-level production.

It is also important to consider that the categorisation of knowledge as implicit depends on how implicitness itself is defined, as well as the measures used to elicit it

(DeKeyser, 2003; Rebuschat & Williams, 2012). Implicit knowledge is commonly characterised by a lack of awareness (DeKeyser, 2003; R. Ellis, 2005, 2009b; Rebuschat & Williams, 2012). As such, online (e.g. think aloud protocols, subjective measures such as confidence ratings and source attributions) and offline (e.g. retrospective verbal reports) measures of awareness have been advocated as means of determining whether a learner is aware of the target grammatical structure when completing a given task and, therefore, whether their performance may be indicative of explicit or implicit knowledge (Hama & Leow, 2010; Leow, 2001; Marsden & Chen, 2011; Rebuschat & Williams, 2012; Williams, 2005). The present study did not include any measures of awareness within the outcomes measures. Nevertheless, the Sentence Reconstruction task demonstrated that, at post-test, the learners were aware of and able to talk about the target grammatical feature and its function within the input. Consequently, in line with a definition of implicit knowledge as knowledge which learners are "generally not aware of possessing" (Rebuschat & Williams, 2012, p. 832), the participants' knowledge of the target feature cannot be labelled conclusively as implicit (DeKeyser, 2003).

6.3.3.3 Automatized explicit knowledge?

Due to the limitations identified in the previous sub-section, it is not possible to state definitively that the learners in the present study had developed a more implicit form of knowledge of the target feature. Nevertheless, the range of outcome measures utilised provided evidence that the learners were not only able to correctly comprehend and produce the target feature on untimed, clozed-response, written activities, but also on more time- and communicatively-pressured oral tasks. It may be, then, that the learners had begun to automatize their explicit knowledge of the target feature, resulting in a proceduralised and, eventually, automatized form of knowledge which was accessible under the time pressure afforded in the oral activities at post-test.

Such a proposal would be in accordance with skill learning theories of language learning, which posit that learning occurs as a result of explicit, declarative knowledge becoming proceduralised and eventually automatized through repeated engagement with the target feature via meaningful activities (DeKeyser, 1997, 2007; DeKeyser & Criado, 2012). By its nature automatized explicit knowledge can be accessible under time and communicative pressure, as is evidenced in previous studies by a reduction in reaction times and error rates (DeKeyser, 1997, 1998). In addition, automatized knowledge exists as a separate knowledge source to the learners' declarative verbalisable knowledge (DeKeyser, 2007) and may not necessarily be verbalisable or available for conscious reflection

(DeKeyser, 1997; Marsden & Chen, 2011). Therefore this interpretation is also in line with the delayed post-test PCA results. Although the learners' ability to verbalise their explicit knowledge had declined at delayed post-test, the learners were able to rely on their separate automatized explicit knowledge when completing the written and oral tasks at this time point, resulting in the six outcome measures loading as one component. An explanation based on the proposed development of automatized explicit knowledge, then, would be compatible with the findings of the present study.

6.3.4 Summary of discussion for RQ2

In answer to RQ2, the present study has provided evidence to suggest that, at post-test, the learners had developed explicit, verbalisable knowledge of the target grammatical feature. In contrast, analysis of the oral tasks indicated that, following the TE-FMC and TE-F interventions, the learners had also developed a type of knowledge that was accessible when performing tasks under a greater level of time and communicative pressure. Notably, based on the methodological considerations set out above, it is not possible to conclude that this constituted a more implicit form of knowledge. Rather the findings of the present study seem to indicate that the learners were, in fact, relying on a more automatized form of explicit knowledge. This finding is also reflected in the learners' continued ceiling level performance on the written and oral tasks at delayed post-test, despite the observed decline in the learners' ability to articulate their explicit, declarative knowledge.

6.4 RQ3 Following explicit information, is intentional practice in attending to the target form-meaning connection more beneficial than intentional practice in attending to the target grammatical form only?

6.4.1 Equivalence of the interventions

One notable outcome of the present study was the finding that both the TE-FMC and TE-F interventions resulted in equivalent, sustained[81] learning gains across all of the outcome measures at post- and delayed post-test. Improvement of the TE-FMC group was in line with previous research which has consistently observed substantial learning gains following referential, structured input activities (i.e. TE-FMC activities as in the present study), both when practice is preceded by explicit information (e.g. Agiasophiti, 2013; Benati, 2001; Benati, 2005; Cheng, 2004; Marsden, 2006; VanPatten & Cadierno, 1993; VanPatten &

[81] with the exception of the Sentence Reconstruction task

Wong, 2004) and when no explicit information is given (e.g. Benati, 2004; Fernandez, 2008; Marsden & Chen, 2011; Sanz, 2004; Sanz & Morgan-Short, 2004; Stafford et al., 2012; VanPatten & Oikkenon, 1996). Based on the findings of such studies it has been argued that the observed effectiveness of PI is due to task-essential attention to the relevant FMC, which is enforced via the referential activities and which, subsequently, leads to improvement in the learners' processing of target language input (Farley, 2004a; Marsden, 2006; Marsden & Chen, 2011; Sanz & Morgan-Short, 2004; Stafford et al., 2012; VanPatten & Oikkenon, 1996). The substantial learning gains made by the TE-FMC group were in line with this argument.

Crucially, the present study has demonstrated that providing learners with input-based instruction, consisting of explicit information plus enriched input activities in which attention to the grammatical form only (i.e. not the FMC) was task essential, was equally as effective as TE-FMC in bringing about improvement in the TE-F learners' comprehension and production of the target feature. It is important, therefore, to consider possible explanations as to why the two interventions were equally effective in helping the learners to circumvent their overreliance on word order which, as evidenced by the pre-test scores, had dominated the learners' interpretation of target language input prior to the intervention.

6.4.2 Accounting for the TE-F group's improvement

6.4.2.1 Encouraging engagement with the target grammatical form

The comparison drawn between the TE-FMC and TE-F interventions builds on previous research which has compared the PI package (or components thereof[82]) with an equivalent form of input-based instruction, namely enriched input (Marsden, 2006; Marsden & Chen, 2011). Neither Marsden (class 1; 2006) nor Marsden and Chen (2011) observed learning gains following enriched input; findings which provided support for the proposal that simply exposing learners to a grammatical form, without pushing them to attend to it in some way, would not result in learning (DeKeyser, 1995; Marsden, 2006; Marsden & Chen, 2011; Schmidt, 1990; VanPatten, 1996, 2004a). Similarly, previous research has demonstrated only minimal effects for input-based instruction which has utilised some form of visual input enhancement (e.g. Agiasophiti, 2013; S.-K. Lee & Huang, 2008; Reinders & Ellis, 2009; J. White, 1998). In contrast, the task-essential attention to form included in the TE-F activities

[82] Marsden (2006) utilised full PI (explicit information plus referential and affective activities). Marsden and Chen (2011) did not include explicit information and compared three interventions: Referential plus Affective activities, Referential activities only, Affective activities only

(clicking on or circling instances of ***der*** and ***den*** within the sentence stimuli) might have been more effective than enriched or enhanced input alone in pushing the learners to attend to the target grammatical form and, subsequently, bringing about substantial learning gains. This provides support for Svalberg's (2012) proposal that incorporating tasks which push learners to engage actively with the target form may serve to enhance the otherwise minimal impact observed for enhanced and enriched input. Indeed, whilst input enhancement techniques are designed to increase learners' awareness of the target feature by manipulating the input in some way (S.-K. Lee & Huang, 2008; Sharwood Smith, 1993; Simard, 2009), there is no guarantee that the learner will actually notice the target feature, even when instruction includes the provision of explicit information (e.g. Marsden, 2006). Therefore the 'noticing' task included in the TE-F activities may have served to orientate the learners' attention repeatedly and explicitly onto the target grammatical form, thereby ensuring that the learners had noticed its presence within the input and could, consequently, benefit from the increased exposure afforded through the TE-F activities. Such a proposal is in line with the view that learning can occur as a result of the initial explicit recognition of a given feature within the input which, in turn, facilitates implicit learning during subsequent exposure to repeated instances within the input (e.g. Implicit Tallying Hypothesis, N. Ellis, 2002; Noticing Hypothesis, Schmidt, 1990).

6.4.2.2 Provision of explicit information

It is also important to consider the potential impact of providing explicit information as part of the TE-FMC and TE-F interventions. Numerous studies have cited evidence that the provision, or absence, of explicit information did not mediate the effectiveness of structured input activities, with no greater learning gains being observed for learners who had received full PI (explicit information plus structured input activities) versus those who had completed structured input activities only (e.g. Benati, 2004; Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Sanz, 2004; Sanz & Morgan-Short, 2004; Stafford et al., 2012; VanPatten & Oikkenon, 1996; Wong, 2004b). Although such findings lend themselves to the proposal that structured input activities are the "*necessary and perhaps sufficient* component of PI" (Farley, 2004a, p. 238), a number of studies (e.g. Culman et al., 2009; Farley, 2004a; Henry et al., 2009; VanPatten & Borst, 2012) have demonstrated the facilitative role that explicit information (either pre-practice information or metalinguistic feedback) can play in promoting the learning of a given grammatical feature. A series of studies demonstrated that providing explicit information served to speed up the rate at which learners began to correctly process the target grammatical feature[83] (Culman et al., 2009; Henry et al., 2009; VanPatten & Borst, 2012). Indeed, in the present study, examination of the TE-FMC learners' scores on the intervention activities suggested that many of the learners were correctly processing the target feature from the first intervention session (Appendix 26). In addition, Wong (2004b) proposed that explicit information may be of particular benefit for instructional approaches which do not make attention to the target FMC task essential, since enriching or enhancing the input may not be sufficient to result in learners attending to the target FMC.

Based on the above observations, then, it could be argued that it was the explicit information provided to both the TE-FMC and TE-F groups (at three points throughout the respective interventions), which accounted for the TE-F group making equivalent learning gains to those of the TE-FMC group. Notably, both groups received exactly the same pre-practice explicit information, which included not only information about the target grammatical feature, but also information relating to the processing problem (i.e. an overreliance on word order), which English learners of L2 German often encounter. Therefore, as proposed by Wong (2004b), the explicit information may have served to focus the TE-F learners' attention on the target FMC, even though it was not required for the activities they were completing. Additionally, such an observation is in line with previous research which has suggested that more explicit forms of input-based instruction (e.g. including the provision of grammatical explanation) may be more effective in inducing noticing than those which provide, for example, enriched or enhanced input only (Reinders & Ellis, 2009; J. White, 1998, 2008). In spite of this suggestion, however, previous research has also demonstrated that providing explicit information alongside enriched input does not necessarily result in learning. The Enriched Input group (in Class 1) in Marsden's (2006) study, for example, exhibited no learning gains, despite having received the same explicit information about the target feature as the PI group. Furthermore, it is prudent to note, again, that a number of studies have also demonstrated that the provision of explicit information alone is not sufficient to result in learning gains (e.g. Benati, 2004; Fernandez, 2008; Marsden & Chen, 2011; Sanz, 2004; Sanz & Morgan-Short, 2004; Stafford et al., 2012; VanPatten & Oikkenon, 1996)

With these observations in mind, then, it is not possible to conclude that the learning gains of the TE-F group can be solely attributed to the pre-practice explicit information. Indeed, it seems that the success of the TE-F intervention may be embedded in the interplay between the explicit information and the 'noticing' task included in the TE-F

[83] Nominative-accusative case marking on masculine definite articles in German

activities. The explicit information provided the learners (in both groups) with portable, declarative knowledge of the target feature and its function within the input (Culman et al., 2009), and the 'noticing' tasks within the TE-F activities kept the learners focussed on the target grammatical form. This combination may have resulted in the TE-F learners attending to the target FMC within the input, despite it not being required for the activities themselves. Consequently, the TE-F learners' processing of the target feature improved to the same extent as that of the TE-FMC group.

In his discussion of the implications of Skill Acquisition Theory for instruction, DeKeyser (2007) observed that the "whole sequence of proceduralization and automatization cannot get started if the right conditions for proceduralization are not present (the declarative knowledge required by the task at hand, and a task set-up that allows for use of that declarative knowledge)" (p. 100). Based on this observation it could be argued that, along with the TE-FMC intervention, the TE-F instruction provided a suitable environment in which learning (i.e. proceduralization and automatization of knowledge) could take place. Learners were provided with the necessary declarative knowledge, which they were then able to apply in tasks which were meaning focused, yet ensured the learners were noticing and, subsequently, processing the target grammatical form. It is also worth noting that the TE-F intervention may have promoted vocabulary learning to a greater extent than the TE-FMC intervention, given the meaning-focused nature of the TE-F activities.

One consideration worthy of note is whether the learners would still have been successful in making the correct FMC had they not received the explicit information. Comparing the above findings with instruction consisting of TE-F activities only (i.e. without the provision of explicit information) would be useful in further elucidating the relationship between these two components of the TE-F intervention and would shed further light on the source (i.e. explicit information, 'noticing' tasks, or both) of the observed improvement following the TE-F intervention.

6.4.3 An exception: Sentence Repetition task

There was one exception to the equivalent learning gains made by the TE-FMC and TE-F groups across the outcome measures. On the oral Sentence Repetition task a certain amount of divergence was observed between the two groups at delayed post-test. First, an approaching significant difference was observed between the TE-FMC and TE-F learners' overall scores at delayed post-test. Secondly, analysis of the Got It and Not Got It sub-groups revealed that the percentage of TE-FMC learners classified as Got It at delayed post-test had increased from post-test and was substantially higher than for the TE-F group (TE-FMC, 51%; TE-F, 29%). Both of these findings indicated that the TE-FMC learners were outperforming the TE-F learners at delayed post-test on this task. Notably, examination of the learners' article use revealed that this divergence between the two groups at delayed post-test was due to an increase in the TE-F learners' tendency to overgeneralise the target feature ***den*** to test items which required the subject article. This tendency was also observed on the Act-Out Production task. In contrast, whilst the TE-FMC learners were overgeneralising their use of ***den*** at post-test, this had decreased by delayed post-test.

Elicited imitation tasks, such as the Sentence Repetition task utilised in the present study, constitute a test of grammatical proficiency since it is argued that a learner will only be able to imitate correctly those grammatical features within the stimuli which are already part of the learners' interlanguage system (Erlam, 2006, 2009). Notably, both the TE-FMC and TE-F learners showed significant improvement in their production of the target feature ***den*** for the object condition in both SVO and OVS sentences, indicating that the target feature had become assimilated into the learners' interlanguage to a certain extent. Furthermore, the findings suggested that, by delayed post-test, a majority of the TE-FMC learners had moved passed the learning stage which is characterised by an increase in errors (i.e. overgeneralisation) and had begun to consistently produce the target feature correctly without overgeneralising to non-obligatory contexts (Kellerman, 1985; Long, 2010; Long & Robinson, 1998). In contrast, the higher incidence of overgeneralisation by the TE-F group suggested that these learners were at an earlier stage in the learning process than the TE-FMC learners; a stage which is characterised by 'backsliding', i.e. an increase in errors, in overall performance ability (Kellerman, 1985; Long, 2010; Long & Robinson, 1998).

It could be argued, then, that the divergence between the two groups was reflective of a difference in the speed at which the TE-FMC versus TE-F learners were acquiring the target feature. In turn, this difference could be attributed to the nature of the two types of intervention. The TE-FMC intervention consisted of explicit information plus activities in which attention to the target FMC was repeated, explicit and task-essential. As such the TE-FMC intervention could be described as a type of explicit, deductive instruction (DeKeyser, 2003; DeKeyser et al., 2002; R. Ellis, 2006). In contrast, the TE-F intervention constituted a more inductive approach. The learners were provided with the same pre-practice explicit information as the TE-FMC group; however the activities themselves only made attention to the target grammatical form, but not its function, task-essential. Therefore any attention to the target FMC which had occurred could be described as induced, since it had occurred independently of the task requirements. In line with this distinction between the two interventions, previous research has demonstrated that more explicit, deductive forms of instruction tend to result in greater learning gains than more implicit or inductive approaches (DeKeyser, 1995; Norris & Ortega, 2000; Robinson, 1997), although notably such findings have tended to be based on learners' performance on outcome measures which favour the use of more explicit knowledge (Doughty, 2003, 2004). It has also been argued that utilising a more explicit form of instruction may lead to the learner attending to and beginning to correctly process a given grammatical feature sooner than a more implicit approach (e.g. DeKeyser et al., 2002; Farley, 2004a; Henry et al., 2009; Marsden, 2006; Marsden & Chen, 2011; VanPatten & Borst, 2012), since implicit learning necessarily requires substantial and repeated exposure to instances of the target feature within the input (R. Ellis, 2002). Consequently, whilst both the TE-FMC and TE-F interventions constituted forms of explicit instruction, the more explicit, deductive nature of the TE-FMC intervention may account for why these learners appeared to have assimilated the target feature into their interlanguage in more target-like ways sooner than the TE-F learners, at least on this particular task.

6.4.4 Individual learner differences: variation in overall test performance

Many, if not a majority of, effect of instruction studies have tended to utilise macro-level analysis relating to the *average* performance of a group(s) of learners, in order to determine the overall effectiveness of a given instructional technique (Larsen-Freeman, 2006). Likewise, the results discussed thus far were primarily based on a comparison of the *average* performance of the TE-FMC, TE-F and Control groups. Such analysis can be useful in determining whether any overall change has occurred in a groups' *average* performance between two (or more) time points or groups of learners. Nevertheless it is important to note, as Larsen-Freeman (2006) argued, that "group averages can conceal a great deal of variability" (p. 598). Such variability is evident in the large standard deviations, which are often reported for learners' post-test (and delayed post-test) performance; yet, micro-level analysis into the factors underlying such variation rarely occurs within the literature (Larsen-Freeman, 2006).

With regard to the present study, despite the fact that no difference was observed between the TE-FMC and TE-F groups as a whole, a substantial increase in the variation (standard deviation) within each group's score was observed at both post- and delayed post-test. Consequently, 'micro-level' analysis was carried out with the intention of illuminating the nature of this increased variation. A distinctive bimodal distribution was observed in the scores of the TE-FMC and TE-F groups on each of the respective outcome measures at post- (Figures 5.1a, 5.1b, 5.1c) and delayed post-test (Appendix 20). Furthermore, division of the learners (within the TE-FMC and TE-F groups, respectively) into sub-groups yielded striking results. Across all outcome measures there was a sub-group of learners (Got It) who were performing at ceiling level and a second sub-group of learners (Not Got It) who continued to perform at chance level following the interventions. Notably this sub-group analysis appears to have produced results contrary to predictions made based on the statistical phenomenon regression to the mean (RTM); namely that the scores of those learners who are at the extreme ends of the distribution at pre-test (i.e. received the lowest and highest marks) are likely to move towards the mean, or middle of the distribution, upon re-testing (i.e. at post- and/or delayed post-test) (Campbell & Stanley, 1966; Marsden & Torgerson, 2012). Contrary to this prediction a consistently bimodal distribution was yielded from the TE-FMC and TE-F groups' data at post- and delayed post-test, with scores tending to be clustered at the two extreme ends of the distribution (i.e. ceiling- and chance-level). It is important to note that within-group analysis of the distribution of learners in the sub-groups at post- and delayed post-test (McNemar test) revealed that there was little movement of learners between sub-groups for any of the tasks, i.e. those learners who were categorised as Got It (or Not Got It) at post-test remained so at delayed post-test. The two exceptions to this were the Sentence Repetition task, for which a number of TE-F learners were found to move down to the Not Got It sub-group at delayed post-test (section 5.2.6), and the Sentence Reconstruction task for which there was an increase in the number of learners within the Not Got It sub-group at delayed post-test (section 5.2.7).

Crucially, then, the overall significant improvement in the two groups' performance across the tasks at post- and delayed post-test can be accounted for by the ceiling level performance of a sub-group of learners within each group. Furthermore, despite the fact that within the TE-FMC and TE-F groups, respectively, the learners had received the same instruction and completed the same activities, for a sub-group of learners (Not Got It) the intervention they had received was not effective in improving their comprehension and / or production of the target feature. It is also important to note that the proportion of learners classified as Got It versus Not Got It varied between tasks, with a higher percentage of learners performing at ceiling level on some tasks (e.g. Sentence Matching), than on others (e.g. Act-Out Comprehension). Larsen-Freeman (2006) observed a similar divergence between participants in her study which, over the course of six months, tracked the development of complexity, accuracy and fluency in the oral and written ability of five

Chinese learners of L2 English. Larsen-Freeman (2006) found that "although … the learners were exposed to similar instructional procedures during the course of [the] study, they actually exhibit[ed] diverging patterns of development" (p. 601), as well as fluctuations in each individual's performance on the specific measures used.

These findings are in line with the view that the language development of individuals within a given language environment (e.g. classroom) is a highly dynamic process characterised by both intra- and inter-individual variation (Larsen-Freeman, 1997, 2006, 2014). From a complexity theory perspective it has been argued that, whilst a "grand sweep" view might posit that learners follow a similar developmental path, there are likely to be qualitative differences between individuals (de Bot, Lowie, & Verspoor, 2007; Larsen-Freeman, 2006) due to variations in the way in which they interact with, attend to and respond to different elements of the environment. Similarly Pienemann (1989) proposed that, although learners tend to move through similar developmental stages in the acquisition of grammar, they are likely to do so at different rates due to variation, for example, in each individual's processing capacity. Furthermore, Marchman and Thal (2005) observed that "individual differences are a natural consequence of language learning" (p. 150). Not only is the developing system constrained by intrinsic factors, such as attention, memory and motivation, but such factors are constantly interacting with external factors, such as the context in which the learning takes place (de Bot et al., 2007). This multi-faceted interaction in turn gives rise to the variability, which is inherent to the development of language within any one individual (de Bot et al., 2007; Larsen-Freeman, 2006, 2014; Marchman & Thal, 2005). Such differences between individual learners are likely to have contributed to the variation observed in the performance of both the TE-FMC and TE-F groups at post- and delayed post-test and the finding that some learners failed to make any improvement following their respective intervention. The role of individual differences, specifically the subcomponent of language learning aptitude grammatical sensitivity (as measured by one part of the MLAT-E), will be discussed in more detail in the following section.

6.4.5 Role of grammatical sensitivity

Fine-grained analysis of the TE-FMC and TE-F learners' performance across the outcome measures highlighted an important consideration in effect of instruction research more generally, as to the way in which individuals respond to a particular teaching approach and the significance of individual differences in mediating success in foreign language learning (Dörnyei & Skehan, 2003). Grammatical sensitivity constitutes one such factor, which has been identified as potentially influencing how a learner responds to a given instructional approach, in particular form-focussed instruction (de Graaf, 1997; Robinson, 1995; VanPatten & Borst, 2012). In line with this proposal, grammatical sensitivity was found to be a significant covariate for both the TE-FMC and TE-F groups across all of the outcome measures, indicating that it was one factor which affected the learners' performance when completing the tasks at post- and delayed post-test. Similarly, these findings are in line with those of VanPatten and Borst (2012), who observed a trend towards grammatical sensitivity being a significant covariate, which exerted a marginal effect on the participants' performance on referential activities following explicit information. Notably, DeKeyser (2000) found that analytical ability (i.e. grammatical sensitivity) was not a significant predictor of success for child second language learners and proposed that this was because children are able to rely on more implicit learning mechanisms. However, DeKeyser's (2000) findings related to learning which had taken place within a naturalistic setting. In contrast, in an instructed setting in which young learners are required to rely on more explicit problem solving abilities, due to limited exposure to target language input, it follows that grammatical sensitivity would have a mediating effect, as was observed in the present study.

It is also important to note, however, that (in the present study) although regression analysis revealed a significant relationship between grammatical sensitivity and certain outcome measures, the r^2 values yielded were consistently small (< .3). VanPatten and Borst (2012) similarly observed only a weak correlation between grammatical sensitivity and the +EI participants' performance[84], with their regression analysis yielding an r^2 value of only 0.191. Therefore, grammatical sensitivity (as measured by one part of the MLAT-E) may have been only one of a number of factors (e.g. working memory, limited attentional resources, general language proficiency, the instruction) which may have influenced the participants' performance at post- and delayed post-test.

6.4.5.1 Relationship between grammatical sensitivity and written task performance

Across all of the analyses (ANCOVA, correlation, regression) a relationship was consistently observed between grammatical sensitivity and the learners' performance on the written measures. These tasks were designed (written, untimed) specifically with the intention of eliciting the learners' explicit knowledge of the target feature. Consequently, the observed correlation was in line with the proposal that tests of grammatical sensitivity (such as the

[84] One treatment group received explicit information (+EI) and one did not (-EI)

MLAT-E Matching Words task) “measure abilities that control access to the learned, but not the acquired, system” (Robinson, 1997, p. 54) and with previous studies which have demonstrated a significant relationship between grammatical sensitivity and explicit learning (e.g. DeKeyser, 2000; Robinson, 1997; VanPatten & Borst, 2012), as well as metalinguistic ability (e.g. Alderson et al., 1997; Ranta, 2002; Roehr, 2006). Overall, the strongest relationship was observed between the learners’ grammatical sensitivity and their performance on the Sentence Matching (written comprehension) task. This is likely to be, in part, a reflection of the nature of the two activities and the fact that both required the learners to perform a similar task; namely discerning the function of key words within the written sentence-level stimuli. Further grammatical sensitivity has been described as being passive in nature, in that it refers to the ability to analyse and detect patterns within the input (Dörnyei & Skehan, 2003; Skehan, 2002), an ability which was required in order to complete the Sentence Matching task successfully.

6.4.5.2 Relationship between grammatical sensitivity and oral task performance

The proposed relationship between grammatical sensitivity and the ability to access explicit knowledge also potentially sheds light on why the analysis produced less consistent results for the oral tasks. In contrast to the written tasks, grammatical sensitivity was not found to be a significant explanatory variable (based on the regression analysis) for either the Act-Out Comprehension or Act-Out Production tasks. This may be due to the fact that these tasks exerted a much greater time and communicative pressure on the learners, thereby restricting the learners’ ability to access and rely on their explicit knowledge.

Contrary to the Act-Out Comprehension and Production tasks, however, significant correlation was observed between the learners’ grammatical sensitivity and their performance on the oral Sentence Repetition task. Similarly Harley and Hart (2002) observed a significant relationship between late immersion learners’ analytic ability and their performance on a Sentence Repetition task similar to that of the present study. Arguably this relationship is a perhaps somewhat surprising result given the nature of the two tasks in question. The Matching Words task is arguably a test of learners’ more explicit grammatical awareness (R. Ellis, 2004; Krashen, 1981; Robinson, 1997), whereas Elicited Imitation tasks, such as the Sentence Repetition task in the present study, are thought to tap into learners’ underlying grammatical competence (Erlam, 2006, 2009). Based on these assumptions a correlation between the two tasks was not anticipated. Notably, however, as acknowledged by R. Ellis (2009c), relatively little research has addressed the question of whether sub-components of aptitude, such as grammatical sensitivity and language analytical ability, can also influence the development of more implicit knowledge. Indeed in one such study, Ranta (2002) demonstrated that language analytic ability played a role in the development of L2 (English) proficiency[85] for young learners learning within a communicative, immersion classroom environment. In this context very little, if any, emphasis was placed on form-focussed instruction; rather the learners' language development was more akin to the process Krashen (1982) terms 'acquisition'. Yet Ranta (2002) observed an association between the learners' language analytical ability and their performance on a number of L2 proficiency measures (e.g. listening comprehension, aural vocabulary recognition).

It is also important to consider, again, that no one task can provide a pure measure of either explicit or implicit knowledge. Most activities are likely to draw on both knowledge types to a certain extent (R. Ellis, 2005, 2009b). In the present study, then, when completing the Matching Words task the learners may have been relying on intuition to a certain extent. Likewise, as the Sentence Repetition task was untimed, and fluency features such as pauses were not analysed, it is possible that the learners may have been drawing on more explicit knowledge when repeating the sentence stimuli. It is not, therefore, possible to state definitively whether the learners were relying on explicit or implicit knowledge, or both, when performing the respective tasks. Consequently the exact nature of the observed relationship between these two tasks is unknown. Further research into the nature of knowledge derived from the two types of task could help to clarify this relationship.

6.4.5.3 Relationship between grammatical sensitivity and metalinguistic task performance

With regard to the Sentence Reconstruction task the relationship between the learners' metalinguistic knowledge and their grammatical sensitivity was less clear cut. A significant correlation was observed for only the TE-F learners at post-test, yet for both groups at delayed post-test. Furthermore, in the regression analysis grammatical sensitivity was found to be a significant explanatory factor for both groups at delayed post-test only. This is contrary to previous research which has demonstrated a significant relationship between learners' grammatical sensitivity and their L2 metalinguistic ability (e.g. Alderson et al., 1997; Ranta, 2002; Roehr, 2006). Crucially, however, the construct of grammatical sensitivity does not require learners to possess metalinguistic knowledge, i.e. be able to talk

[85] L2 proficiency tests: aural vocabulary recognition, listening comprehension, cloze gap-fill, MEQ listening comprehension, English metalinguistic task, yes/no vocabulary (Ranta, 2002)

about language (VanPatten & Borst, 2012). Yet it was this ability to *talk about* the target grammatical feature which was directly tested in the Explanation element of the Sentence Reconstruction task and may, therefore, account for the lack of consistent relationship between the learners' performance on this task and their grammatical sensitivity at post-test.

6.4.5.4 Relationship between grammatical sensitivity and the TE-F intervention

Both the correlation and regression analyses revealed that the relationship between grammatical sensitivity, i.e. the ability to seek patterns in the input (Carroll, 1973; Dörnyei & Skehan, 2003; Skehan, 2002), and performance on the six outcome measures was stronger for the TE-F group than for the TE-FMC group at post-test. This suggested that grammatical sensitivity may have played a greater role in mediating the TE-F group's performance at post-test, which may have been due to the nature of the intervention activities completed by the two groups. The TE-FMC intervention required the learners to complete activities in which attention to the relevant FMC had been made task-essential. It could be argued that this intervention overcame the learners' grammatical sensitivity through repeatedly and explicitly directing their attention to the FMC. In contrast, the TE-F activities required the learners to attend to the grammatical form only (by clicking on, or circling, the form in the input). Therefore any attention which had been paid to the FMC was above and beyond the task-essential requirements of the activities and, arguably, more susceptible to the learners' own level of grammatical sensitivity. Robinson (1997) observed significant correlations between learners' grammatical sensitivity and their accuracy on a grammaticality judgement task, not only for the more explicit instructed and rule-search conditions, but also the implicit condition. Robinson (1997) proposed that the learners in the implicit condition may have also begun to "consciously analyze, search for and find the rules underlying the presented sentences" (p. 75); a behaviour which is likely to be predicted by the grammatical sensitivity component of aptitude. Similarly Ranta (2002) proposed that those learners with a higher level of language analytical ability are likely to be most successful at inducing the target grammatical rule under more implicit forms of instruction (e.g. input enhancement). With regard to the present study, those TE-F learners with a high-level of grammatical sensitivity may have been more successful at actively attending to the target FMC in the input of their own volition and in addition to the demands of the activities themselves. Consequently grammatical sensitivity may have played a bigger role for this group than for the TE-FMC learners. To unpick the relationship between grammatical sensitivity and the two types of instruction in more detail, future analysis could correlate the grammatical sensitivity scores and test performances of the Got It and Not Got It sub-groups within each intervention group respectively (i.e. investigate whether the Not Got It sub-group also scored lower on the grammatical sensitivity task).

6.4.6 Summary of discussion for RQ3

In summary, the answer to RQ3 would appear to be broadly 'No'; following explicit information, intentional practice in attending to the target form-meaning connection was not more beneficial than intentional practice in attending to the target grammatical form only. Overall, both interventions were successful in pushing the learners to attend to the target grammatical form and the relevant FMC, resulting in an improvement in both the TE-FMC and TE-F learners' processing of target language input. Notably, the Sentence Repetition task proved to be one exception. It was proposed that the more explicit nature of the TE-FMC intervention may have resulted in the target feature being more fully integrated into the TE-FMC learners' interlanguage sooner than for the TE-F learners, resulting in the higher incidence of overgeneralisation by the TE-F learners at delayed post-test.

Finally, whilst no difference was observed in terms of the overall effectiveness of the two interventions, analysis of the Got It and Not Got It sub-groups demonstrated the differential impact which instruction had for individual learners. Grammatical sensitivity was found to be one factor which mediated the impact of instruction to a certain extent. Furthermore, grammatical sensitivity was found to play a larger role in the TE-F groups' performance, which it is argued may have been due to the more inductive nature of the intervention.

Chapter 7: Conclusion

7.1 Summary of the study

This thesis has presented the findings of a classroom-based, experimental study investigating the effectiveness of explicit grammar instruction for young foreign language learners. The participants were 138 L1 English learners (aged 9 to 11) of L2 German from three primary schools (seven classes) in Yorkshire, who had been learning German in school (taught by the researcher) for approximately one year prior to the study.

The study sought to compare the effectiveness of two types of input-based grammar instruction; TE-FMC and TE-F. TE-FMC utilised two components of the instructional approach identified as Processing Instruction, namely explicit information coupled with activities in which attention to the target FMC was task-essential. In contrast, whilst the TE-F intervention utilised the same explicit information as TE-FMC, the TE-F activities made attention to the target grammatical form only (i.e. not its function) task-essential. The interventions were administered in weekly sessions over a period of five weeks, giving a total duration of 4 hours. In addition a non-active Control group was utilised in order to control for any potential test effect influencing the learners' performance.

The target grammatical feature was accusative definite article case marking (***den***) for masculine nouns in German. It was hypothesised that L1 English learners of German would overlook this grammatical feature, relying instead on word order in order to assign grammatical roles within the L2 input. The participants were tested at three points throughout the study; pre-test (week 1), post-test (week 7) and delayed post-test (week 16). Six outcome measures were utilised in order determine the effectiveness of the two interventions. The outcome measures constituted both written and oral tests of comprehension and production, as well as a metalinguistic task designed to test learners' ability to verbalise the target grammatical rule. The written and metalinguistic tasks were designed to tap into more explicit knowledge, whereas the oral tasks, completed one-to-one with the researcher, were designed to exert a greater level of time and communicative pressure and therefore tap into more automatized and / or implicit knowledge.

7.2 Summary of the findings

The present study has provided substantial evidence that explicit grammar instruction can be effective in improving young learners' processing of definite article case-marking in L2 German. As a result of their respective interventions, the findings suggested that both the TE-FMC and TE-F learners were attending not only to the target grammatical form but also its function within the input, and as such had overcome their reliance on word order when interpreting sentences in the target language. These findings would therefore support the proposal that instruction can serve to push learners away from "less-than optimal" processing strategies; thereby optimising learners' processing of target language input and providing richer intake for the developing system (VanPatten, 2002, 2007; Wong, 2004a). Accordingly the TE-FMC and TE-F learners' performance across the six outcome measures at post- and delayed post-test demonstrated that this improved processing, brought about as a result of the input-based instruction, had not only improved their comprehension, but also their production of the target grammatical feature. These findings were in line with previous research which has demonstrated a similar benefit of input-based instruction for younger learners (e.g. Harley, 1998; Laval, 2013; Mavrantoni & Benati, 2013), older learners (e.g. Agiasophiti, 2013; Benati, 2005; Marsden, 2006; Marsden & Chen, 2011), and adults (e.g. Benati, 2001; Benati, 2004; Cheng, 2004; Stafford et al., 2012; VanPatten & Cadierno, 1993; VanPatten & Wong, 2004). The answer to the first research question, then, is 'yes'; explicit grammar instruction can improve young learners' comprehension and production of the target grammatical feature.

With regard to the type of knowledge developed following instruction (RQ2), the TE-FMC and TE-F learners' post-test performance on the untimed, written tasks indicated that they had developed explicit knowledge of the target feature. Further, on the metalinguistic task, many of the learners were able to verbalise this knowledge utilising the appropriate metalinguistic terminology. Notably, however, by delayed post-test the learners' ability to articulate this knowledge had decayed. The learners' performance on the oral measures, as well as the observed decrease in their metalinguistic task performance at delayed post-test, suggested that the learners may have begun to proceduralise, and to a certain extent automatize, their explicit, declarative knowledge. These findings therefore provide evidence in support of the proposal that explicit instruction, and the ensuing explicit knowledge, can play a role in L2 learning (DeKeyser, 2007; DeKeyser et al., 2002; DeKeyser & Sokalski, 1996; N. Ellis, 2005; R. Ellis, 2006; Schmidt, 1990), resulting in the development of knowledge which is accessible in comprehension and production tasks performed under time and communicative pressure.

With regard to the comparison of the two instructional approaches (RQ3), both the TE-FMC and TE-F interventions proved to be equally effective in improving the learners' overall performance. A small amount of divergence was observed on the oral Sentence Repetition task, indicating that the more explicit, deductive nature of the TE-FMC intervention may have resulted in the target feature being assimilated into the learners' interlanguage more fully than for the TE-F learners. Nevertheless, despite this small amount of variation, both groups were found to make statistically equivalent gains across all of the outcome measures, and further had sustained those gains when tested at delayed post-test. The observed equivalence of the two interventions therefore suggested that instruction, which makes attention to the grammatical form only task-essential, can also be successful in pushing learners to attend to the target grammatical feature and its function within the input. The equivalence of the learning gains made by the two groups suggested that as a result of the explicit information, the 'noticing' task, or a combination of the two, the TE-F learners were also attending to the relevant FMC despite the fact that it was not required for successful completion of the activities. Indeed Wong (2004b) claimed that "the best kind of intervention may be one in which input is structured so that learners can perceive and parse L2 stimuli more effectively" (p. 198); both the TE-FMC and TE-F interventions appear to have been successful in doing so.

Finally the present study found that instruction can have a differential benefit for individual learners. Indeed, whilst an overall significant improvement was observed for both the TE-FMC and TE-F groups following instruction, more fine-grained analysis revealed that this improvement could be accounted for by the ceiling level performance of a sub-group of learners within each group. Further grammatical sensitivity was found to account for some of the variation observed in the learners' performance, although the results of both the ANCOVA and regression analysis suggested that it was only one of a number of factors which may have influenced the learners' performance across the outcome measures. As such it is important to acknowledge the importance of individual differences in mediating the effectiveness of a given instructional approach (Dörnyei & Skehan, 2003; Larsen-Freeman, 2006, 2014; Marchman & Thal, 2005; Skehan, 2002).

7.3 Limitations and future research

It is possible to identify a number of limitations with regard to the present study. Firstly whilst the TE-F intervention was found to produce equivalent learning gains to those of the TE-FMC intervention, it is not clear whether the effectiveness of the TE-F intervention can be attributed to the explicit information, the 'noticing' task, or a combination of the two instructional components. Future research could therefore compare the full TE-F intervention to instruction consisting of the TE-F activities only, in order to clarify whether the 'noticing' tasks alone would be sufficient in pushing learners to notice the target feature and subsequently its FMC when exposed to enriched target language input. Notably it would also be illuminating to carry out such a comparison with the TE-FMC intervention. A series of studies has demonstrated that the referential structured input activities are the "*necessary and perhaps sufficient* component of PI" (Farley, 2004a, p. 238); however such studies have tended to be carried out with slightly older learners (e.g. Marsden & Chen, 2011) than those of the present study, or adults (e.g. Benati, 2004; Sanz & Morgan-Short, 2004; Stafford et al., 2012; VanPatten & Oikkenon, 1996). Therefore it remains to be seen whether younger learners, whose explicit problem-solving abilities are not as highly developed as those of older and adult learners, would still be successful in attending to and inducing the target FMC when their attention has not first been directed there via the explicit information.

A second issue relates to the nature of knowledge developed by the learners following instruction. The written and metalinguistic tasks indicated that the learners had developed functional, and verbalisable (at post-test), explicit knowledge of the target feature. However it was not possible to conclude based on the findings of the present study, whether the learners had also developed implicit knowledge. Indeed it was proposed that repeated engagement with the intervention activities may have served to proceduralise, and to a certain extent automatize, the learners' initial declarative knowledge, as evidenced by their performance on the oral tasks at post-test and across the tasks at delayed post-test. Future studies could therefore incorporate more sensitive measures of implicit knowledge (e.g. timed tasks; spontaneous, discourse-level, oral production tasks; confidence ratings and source attributions; reaction times etc) in order to determine more conclusively the nature of knowledge derived following explicit, input-based instruction.

A third consideration relates to the processing problem which all of the learners were found to encounter at pre-test; namely an overreliance on word order. Whilst the present study confirmed that the participants' interpretation of L2 German was indeed constrained by this processing problem prior to instruction, it was not possible to determine the source of this constraint, i.e. whether this was due to a more universal processing problem, as would be predicted by the FNP (VanPatten, 1996, 2004a, 2007), or whether this was an issue of L1 transfer, as predicted by the Competition Model (MacWhinney, 2001).

Fourthly, the present study illustrated the importance of considering the role of individual differences in mediating the impact of a given instruction approach. The findings suggested that grammatical sensitivity was one factor, which had influenced individual learners' performance to a certain extent, in particular for the TE-F learners. Further analysis could tease out from the data the relationship between grammatical sensitivity and learning under the two instruction conditions, for example by comparing the differential impact of grammatical sensitivity for the Got It and Not Got It sub-groups respectively. Further, future research could consider in more detail the nature of knowledge derived from tests of grammatical sensitivity (e.g. MLAT-E Matching Words task), in order to clarify the nature of the relationship between such tests and tasks requiring the use of more automatized or implicit knowledge.

It is also important to consider that the focus of the instruction in the present study was the comparison of two items (nominative, ***der***, and accusative, ***den***, definite articles) from the wider case marking system in German. Further such a comparison could be considered relatively simple, since both articles carried semantic value (i.e. denoting the subject and object of the sentence respectively) and within the instructional input there was a one-to-one mapping between the grammatical forms and their respective meanings (Spada & Tomita, 2010). Future research could therefore investigate whether younger learners can also benefit from instruction on a wider range of simple, as well as more complex, grammatical features.

Finally, both the TE-FMC and TE-F learners were found to sustain the improvement they had made across the outcome measures at delayed post-test nine weeks after the intervention. This was considerably later than the average delayed post-test in either Norris and Ortega's (2000) (*M* = 4.34, *SD* = 5.02) or Spada and Tomita's (2010) (*M* = 4, *SD* = 3.62) metal-analyses. Nevertheless, it is important to acknowledge the difficulty of assessing the even longer term impact of instruction, particularly within classroom-based research, due to factors such as the learners changing schools (as the Year 6 learners did within a matter of months following the close of the present study) and the difficulty of controlling learners' exposure to the target language (J. White, 2008). As such, whilst the delayed post-test results indicated that the learners were continuing to attend to the target FMC, in preference to the less reliable word order cue, it is not possible to definitively state that the learners would have continued to do so over the longer term.

7.4 Implications for the primary foreign language classroom

The present study has provided substantial evidence that explicit input-based grammar instruction was useful for young learners (aged 9 to 11) of L2 German learning within the input-poor foreign language classroom, as evidenced by the TE-FMC and TE-F groups' significant improvement across a battery of written and oral outcome measures. Both the TE-FMC and TE-F interventions were successful in pushing the young learners to attend to and correctly process the target FMC, which in turn helped the learners to overcome their reliance on word order and rely instead on definite article case-marking (specifically accusative case-marking on masculine articles) when interpreting target language sentences. This finding is in line with a number of recent studies which have demonstrated a positive effect of PI for young learners (Angelovska & Benati, 2013; Laval, 2013; Mavrantoni & Benati, 2013). In addition the present study builds on such work by demonstrating significant, durable gains on a range of different outcome measures.

It is worthy of note that the outcome measures as well as the intervention activities utilised in the present study were specifically designed in order to be appropriate for the age (9 to 11 years old) and L2 ability (beginner) of the learners in question. As such the findings of the study suggest firstly that the picture-based, computerised and paper-and-pen activities were successful in engaging these young learners in form-focussed learning throughout the intervention. Secondly the outcome measures themselves were successful at eliciting language data from the participants. In particular whilst the use of discourse-level tasks was not possible, utilising age-appropriate resources such as soft toys and picture stimuli proved to be fruitful in eliciting more spontaneous, oral language use.

The findings of the present study are also in line with previous research which has demonstrated that alternative types of form-focussed instruction (e.g. input enhancement) can lead to improvements in younger learners' comprehension and production of key grammatical features, although notably such prior research has tended to be carried out with learners of L2 English within the input-rich, immersion classroom setting (e.g. Harley, 1998; Spada & Lightbown, 1993; J. White, 2008; J. White & Ranta, 2002; L. White et al., 1991). Within the immersion setting form-focussed instruction can serve to draw learners' attention to a given grammatical feature, which may then be reinforced and consolidated into the learners' interlanguage through extensive exposure to the target feature within the classroom input (e.g. as predicted by the Implicit Tallying Hypothesis, N. Ellis, 2002). In contrast the limited exposure to target language input afforded within the foreign language classroom context may not necessarily offer the opportunity for such reinforcement of the target feature following instruction (Gass & Selinker, 2008; Muñoz, 2006, 2008b; Philp & Tognini, 2009). Nevertheless the present study has demonstrated that with relatively little input (4 hours over the five week intervention) the learners made substantial progress. Indeed the TE-FMC learners' scores from the intervention activities demonstrated that at least some of the learners had begun to correctly process the target feature as early as during the first intervention session (Appendix 26). It is important to acknowledge, that during the initial intervention session(s) it was likely that the learners were relying primarily on their explicit, declarative knowledge of the target feature, gained through the pre-practice grammar explanation, when completing the untimed intervention activities (DeKeyser, 2007; DeKeyser et al., 2002). Nevertheless the learners' performance on the oral tasks at post- and delayed post-test suggested that this initially declarative knowledge may have become proceduralised and even automatized to a certain extent, through the opportunities to practice provided during the subsequent weeks of the intervention (DeKeyser, 2007; DeKeyser & Criado, 2012). With regard to the foreign language classroom, and in particular the UK context, language lessons at primary school-level are often restricted to, at most, one hour per week (Cable et al., 2012; Marilyn Hunt et al., 2005; Wade et al., 2009); therefore the time available to spend on individual tasks is necessarily limited. The findings of the present study, however, have provided evidence that a relatively small amount of instruction consisting of a short grammar explanation plus weekly practice provided via meaningful activities can result in substantial learning gains. Even within the limited time available in many foreign language classrooms, short weekly 'bursts' of focussed grammar instruction and activities are likely to be achievable, and as evidenced by the present study can have a beneficial effect.

The findings of the present study also provide support for the claim made by Bouffard and Sarkar (2008) that young learners are "mature enough to attend to form if they are taught how to" (p. 22). Indeed as highlighted by Philp et al. (2008), in middle childhood (i.e. aged 7 to 11) children are becoming more logical in their thinking and further are developing a greater level of metalinguistic awareness. The TE-FMC and TE-F learners' performance on the Sentence Reconstruction task demonstrated that the learners had developed and were able to express their metalinguistic knowledge of the target grammatical form, with many learners employing the appropriate metalinguistic terminology when doing so. Foreign language pedagogy, then, could usefully adapt to these changes in learners' cognitive maturity and reflect the fact that as children's analytical skills develop they are able to benefit from more explicit, form-focussed instruction integrated within an overall communicative language curriculum (Philp et al., 2008).

Notably, however, it is also important to consider the intervening role which individual differences can play in determining how a particular learner responds to instruction. The sub-group (Got It versus Not Got It) analysis demonstrated that there were some learners who did not make any gains following the TE-FMC or TE-F interventions. Further the analysis of the grammatical sensitivity task suggested, in line with previous research (e.g. Harley & Hart, 2002; Ranta, 2002; J. White & Ranta, 2002), that the learners' language analytical ability was one factor which may have mediated the effectiveness of the instruction, particularly for the TE-F group. In addition factors such as learners' limited attentional resources (N. Ellis, 2006; VanPatten, 1996, 2004a) as well as their more general language proficiency are likely to influence the impact of form-focussed instruction. Indeed based on the findings of her study, Harley (1998) observed that learners who were preoccupied with remembering novel vocabulary may not have been able to "devote full attention to the formal aspects that were the intended focus of the activities" (p. 169). As such form-focussed instruction within the foreign language classroom should be tailored to suit not only the age but also the cognitive maturity of the learners in question (Cameron, 2001; Philp et al., 2008).

One final finding worthy of note is that the range of outcome measures utilised in the present study demonstrated the positive effect of both the TE-FMC and TE-F interventions on the learners' comprehension and use of the target feature within all four key skills; namely reading (written, comprehension), writing (written, production), listening (aural, comprehension), and speaking (oral, production). This finding is particularly pertinent given that the current KS2 curriculum for the UK stipulates that language teaching should "enable pupils to **understand and communicate** ideas, facts and feelings **in speech and writing**, focused on familiar and routine matters, using their knowledge of phonology, **grammatical structures** and vocabulary" (emphasis added, DfE, 2013d, p. 2). Therefore instructional approaches, such as the TE-FMC and TE-F interventions utilised in the present study, which led to the learners making significant gains in all four skills simultaneously, are likely to be preferred in a context in which the time available for language teaching is limited.

7.5 Contributions of the study

In conclusion, the present study has made a number of significant contributions to the field of research into explicit instruction and language learning.

Firstly the learners within the present study were younger than those in a majority of previous effect of instruction studies, and those studies, which have been carried out with younger learners, have tended to do so within the immersion classroom context (Harley, 1998; J. White, 2008; L. White et al., 1991). In contrast the present study was conducted within the instructed foreign language classroom environment in the UK. As such the findings of the study are relevant to the debate surrounding the role of explicit learning and instruction in child language learning. Indeed whilst it is commonly thought that young learners learn best implicitly (e.g. DeKeyser, 2003; DeKeyser & Larson-Hall, 2005; Lenneberg, 1967), the present study has demonstrated that within the input-poor foreign language classroom, younger learners can also benefit from more explicit instruction and can make significant gains in comprehension and production as a result of more explicit learning. In addition the findings contribute to the wider discussion regarding the role of explicit knowledge in language learning (e.g. DeKeyser, 2003; N. Ellis, 2005; R. Ellis, 2006; Krashen, 1982; Schmidt, 1990). Following instruction the learners in the present study had not only developed explicit knowledge of the target grammatical feature, but also more automatized knowledge which was accessible under greater time and communicative pressure.

Secondly the present study contributes to research comparing the effectiveness of PI, or components thereof, with alternative forms of input-based instruction. To date only a handful of studies have been found to draw such a comparison (Agiasophiti, 2013; Marsden, 2006; Marsden & Chen, 2011). In addition, building on the findings of Marsden’s studies, the present study drew a novel comparison between instruction in which attention to the target FMC was task-essential, and an alternative form of input-based instruction in which attention to the target grammatical form only was required.

Thirdly, in contrast to a majority of previous classroom-based studies, the present study avoided the use of intact classes. Both the TE-FMC and TE-F groups contained a mix of learners from across four separate classes, thereby minimizing the potential impact of extraneous variables such as history and maturation on the outcomes of the study.

Fourthly, the fine-grained analysis of the TE-FMC and TE-F learners’ performance by sub-group (i.e. Got It versus Not Got It) across the outcome measures makes a unique contribution to research in this area. Very few studies have endeavoured to drill down into and explore the performance of individuals within a given treatment group, despite the increase in variation which is often observed in learners' post-test performance following instruction (Larsen-Freeman, 2006, 2014).

Finally, in contrast to many effect of instruction studies, the present study was carried out within an 'ordinary' classroom, using 'ordinary' classroom resources, by the 'regular' German teacher, during each class's weekly German lesson. Therefore the ecological validity and consequently the implications of the present study for classroom-based foreign language learning cannot be denied. As stated by Hulstijn and de Graaf (1994) when it comes to investigations into the effectiveness of a given instructional approach, "the final proof should always be given in a 'normal' language class" (p. 108).

Appendix 1 Principle 1 of VanPatten's Input Processing theory

Principle1a. The primacy of content words principle.
Learners process content words in the input before anything else

Principle 1b. The lexical preference principle.
Learners will tend to rely on lexical items as opposed to grammatical forms to get meaning when both encode the same semantic information.

Principle 1c. The preference for nonredundancy principle.
Learners are more likely to process nonredundant meaningful grammatical form before they process redundant grammatical forms.

Principle 1d. The meaning-before-nonmeaning principle.
Learners are more likely to process meaningful grammatical forms before nonmeaningful grammatical forms irrespective of redundancy.

Principle 1e. The availability of resources principle.
For learners to process either redundant meaningful grammatical forms or nonmeaningful forms, the processing of overall sentential meaning must not drain available processing resources.

Principle 1f. The sentence location principle.
Learners tend to process items in sentence initial position before those in final position and those in medial position.

(VanPatten, 2004a, p. 14)

Appendix 2 Consent form for head teacher / teacher

KEY INFORMATION AND DECLARATION OF CONSENT

PRINCIPAL RESEARCHER: Rowena Hanan

SUPERVISOR: Dr Emma Marsden

PROGRAMME: PhD in Education (Department of Education, University of York)

PROJECT TITLE: The role of explicit knowledge in Primary school level foreign language learning: Is it useful for language learning and can it be trained?

KEY INFORMATION

WHAT IS THE RESEARCH PROJECT INVESTIGATING?

The focus of this research is to explore how children learn and process a foreign language in Primary school. The aim is to determine the effectiveness of different pedagogical tools for language teaching in Primary school.

WHAT DOES THE STUDY ENTAIL?

PRE-TEACHING PHASE

Prior to the study taking place, the researcher (Rowena Hanan) will be teaching German to the Year 5 and 6 classes in the participant schools. This teaching will cover general topics such as Greetings and Introductions, Numbers, Age, Family, Pets and Animals, Hobbies, and School, as well as topics concerned with the culture and traditions in Germany and German speaking countries. The pre-teaching phase does not form part of the main study.

MAIN STUDY

The study will take place over a period of 8 weeks; the Year 6 class will take part in the study during the spring term and the Year 5 class during the summer term. There will be an intervention phase and activities completed at the beginning and end of the study. These are all intended to give the learners practise of basic German vocabulary and promote language awareness. All activities are specifically designed to be appropriate for pupils aged 9-11.

TIMETABLE OF STUDY

WEEK 1: ACTIVITIES (BEFORE INTERVENTION)

Two reading and writing activities will be completed by the whole class at one time. Three speaking and listening activities will be completed one-to-one with the Researcher, lasting approximately 20-30 minutes in total. The one-to-one activities will be recorded on a Dictaphone and / or video camera.

WEEKS 2 - 7: INTERVENTION (6 WEEKS)

Each of the classes will be randomly split into two groups during this phase. Each group will be taught using exactly the same language content in German, but will receive slightly different activities types. The teaching will be via both computer and paper-based activities, created by the researcher.

WEEKS 8 & 16: ACTIVITIES (IMMEDIATELY FOLLOWING & APPROX. 8 WEEKS AFTER INTERVENTION)

Pupils will again complete the short reading, writing, speaking and listening activities completed in Week 1.

PARTICIPATION

Participation in the study is voluntary. A letter, written by the researcher, will be sent home to the Year 5 and 6 parents prior to the main study commencing, to inform them that their children will be taking part in the research project. Nearer the time the Researcher and class teacher will reach a decision about providing the opportunity for parents to 'opt-out' of the study should they wish to.

CONSENT FORM

I understand that the aim of this project is to explore how foreign languages are learnt within the classroom context.

I understand that pupils will be involved in a variety of reading, listening, speaking and writing activities in German.

I understand that pupils' involvement in the study is voluntary and can therefore be withdrawn at any time.

I understand that the data gathered will be stored anonymously and that no unauthorised person will have access to the data.

I understand that audio files generated through the one-to-one activities may be played at conferences but that in such cases participants will remain anonymous.

DECLARATION OF CONSENT

I have been informed about the aims and procedures involved in this research project and consent to the details of research, detailed above.

______________________	______________________	____________
Name of Headteacher	Signature of Headteacher	Date
______________________	______________________	____________
Name of Researcher	Signature of Researcher	Date

Appendix 3 Letter to parents

THE UNIVERSITY *of York*

DEPARTMENT OF EDUCATION

Heslington, York, YO10 5DD

Email:
reh505@york.ac.yuk

23rd January 2013

Dear Parents,

This term Year 6 will be taking part in a research study, conducted by a researcher from the Department of Education at the University of York.

The study is investigating how children learn foreign languages in the Primary school classroom and will form part of the pupils' regular German teaching. Pupils will be completing short reading, writing, listening and speaking activities in German, designed to give pupils practice of basic German vocabulary, as well as promote language awareness through opportunities to practise talking about language.

Pupils will be taking part in one-to-one and group language learning activities; some of the one-to-one activities will be recorded on a video camera. No unauthorised person will have access to the video recordings, and they will be viewed only by the two researchers named below. Short extracts may also be used in research conference presentations. If you would prefer that these videos are not shown at such events, please inform XXXXXX in writing by XXXXXX.

Participation in the study is voluntary; therefore please contact XXXXXX if you would prefer your child not to take part.

If you have any questions about the study, please do not hesitate to contact me.

Yours sincerely

Rowena Hanan
(Email: reh505@york.ac.uk)
PhD Student, Department of Education, University of York

and Dr Emma Marsden
(Email: emma.marsden@york.ac.uk; tel: 01904 323335)
Senior Lecturer in Second Language Education, Department of Education, University of York

Appendix 4 List of core vocabulary

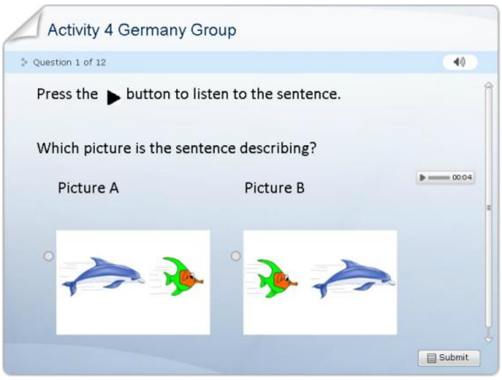
Gender	Nouns: People	Nouns: Animals	Nouns: Objects
Masculine	Mann	Elefant	Brief
	Junge	Hund	Ball
	Vater	Affe	Stuhl
	Sohn	Panda	Tisch
	Onkel	Fisch	Computer
	Opa	Vogel	Bus
	Enkelsohn	Hahn	Traktor
	Schüler	Bulle	Kuchen
	Lehrer	Tiger	Apfel
	Professor	Löwe	Teddybär
	Student	Hamster	
	Farmer	Papagei	
	Clown	Flamingo	
		Bär	
Feminine	Mutter	Kuh	Tür
	Oma	Katze	Banane
	Tante	Maus	Hausaufgabe
	Cousine	Spinne	Uhr
	Tochter	Schlange	Pizza
	Enkeltochter	Biene	
	Frau		
	Lehrerin		
	Fuβballspielerin		
	Tennisspielerin		
Neuter	Mädchen	Kaninchen	Fenster
	Baby	Schwein	Buch
	Kind	Schaf	Eis
		Pferd	Frisbee
		Meer-schweinchen	Fahrrad

Verbs	
begrüßen	to greet
benutzen	to use
besuchen	to visit
erschrecken	to scare
essen	to eat
fotografieren	to photograph
fragen	to ask
füttern	to feed
haben	to have
hören	to hear
kaufen	to buy
kicken	to kick
küssen	to kiss
lesen	to read
lieben	to love
machen	to do / make
milken	to milk (a cow)
öffnen	to open
putzen	to clean
rufen	to call
schlagen	to hit (so-sth)
schreiben	to write
sehen	to see
streicheln	to stroke
treffen	to hit (sth-so)
umarmen	to hug
wäschen	to wash
verfolgen	to chase

Appendix 5 *Which picture?* activity (TE-FMC)

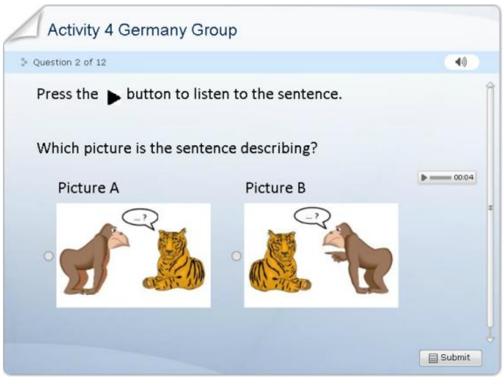
Listening version:

SVO sentence:



Transcript: Der Delfin verfolgt den Fisch. [the-NOM dolphin chases the-ACC fish]

OVS sentence:



Transcript: Den Tiger fragt der Gorilla. [the-ACC tiger asks the-NOM gorilla]

Worksheet version:

SVO sentence:

Der Affe begrüβt den Panda.

[the-NOM monkey greets the-ACC panda]

A

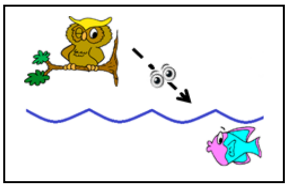
B

OVS sentence:

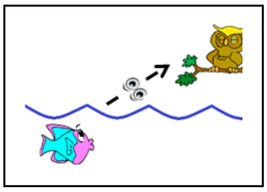
Den Fisch sieht der Vogel.

[the-ACC fish sees the-NOM bird]

A



B



Feedback for worksheet version:

SVO sentence:

Der Affe begrüβt den Panda.

The answer is **A**! The sentence says that the monkey greets the panda.

A

B

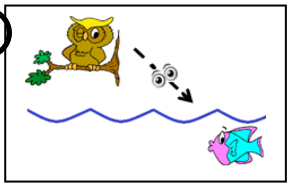
OVS sentence:

Den Fisch sieht der Vogel.

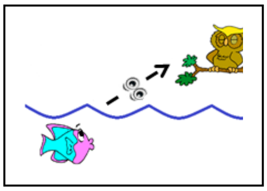
This word tells us that the *Fisch* (fish) is the object and is being seen.

This word tells us that the *Vogel* (bird) is the subject and is doing the seeing.

A



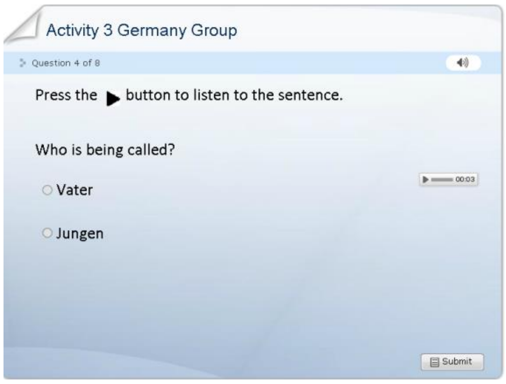
B



Appendix 6 *Who's doing what?* activity (TE-FMC)

Listening version:

SVO sentence:



Transcript: Der Vater ruft den Jungen. [the-NOM father calls the-ACC boy]

OVS sentence:

Transcript: Den Panda küsst der Affe [the-ACC panda kisses the-NOM monkey]

Worksheet version:

SVO sentence:

Der Farmer füttert den Hahn. [the-NOM farmer feeds the-ACC hen]

Who is doing the feeding?	Farmer	Hahn

OVS sentence:

Den Mann verfolgt der Computer. [the-ACC man chases the-NOM computer]

Who is being chased?	Mann	Computer

Feedback for worksheet version:

SVO sentence:

Der Farmer füttert den Hahn.

Who is doing the feeding? Farmer Hahn

This word tells us that the 'Farmer' is the subject of the sentence and is doing the feeding.

OVS sentence:

Den Mann verfolgt der Computer.

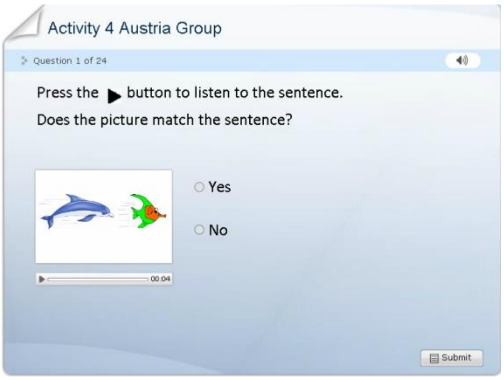
Who is being chased? Mann Computer

This word tells us that the ***Mann*** is the object of the sentence and is being chased.

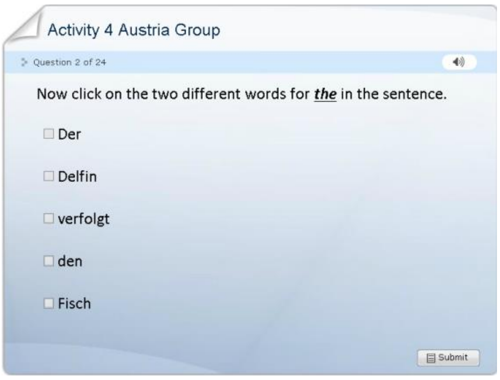
Appendix 7 *Picture matching* activity (TE-F)

Listening version:

Enriched input task (SVO):



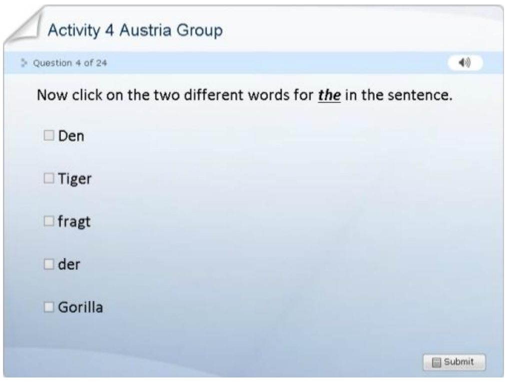
Noticing task:



Transcript: Der Delfin verfolgt den Fisch. [the-NOM dolphin chases the-ACC fish]

Enriched input task (OVS):

Noticing task:



Transcript: Den Tiger fragt der Gorilla. [the-ACC tiger asks the-NOM gorilla]

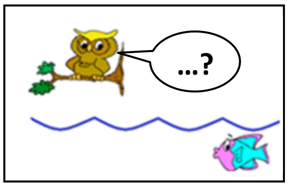
Worksheet version:

SVO sentence:

Der Affe begrüβt den Panda. Yes No
[the-NOM monkey greets the-ACC panda]

OVS sentence:

Den Fisch sieht der Vogel. Yes No
[the-ACC fish sees the-NOM bird]



Feedback for worksheet version:

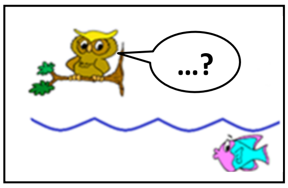
SVO sentence:

Der Affe begrüβt den Panda. Yes No

Yes! The picture does match the sentence. The sentence says that the monkey ***greets*** the panda and that is what we can see in the picture.

OVS sentence:

Den Fisch sieht der Vogel. Yes No



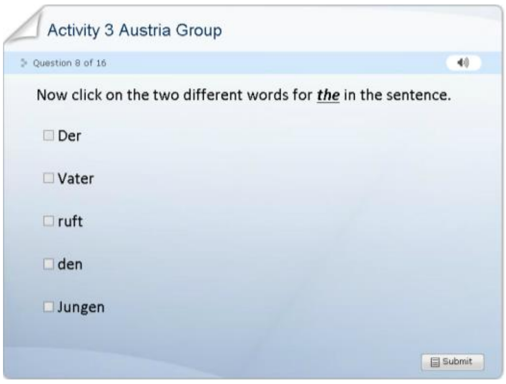
No! The sentence says that the bird ***sees*** the fish. But in the picture the bird is ***asking*** the fish.

Appendix 8 *Sensible or silly?* activity (TE-F)

Listening version:

Enriched input task (SVO):

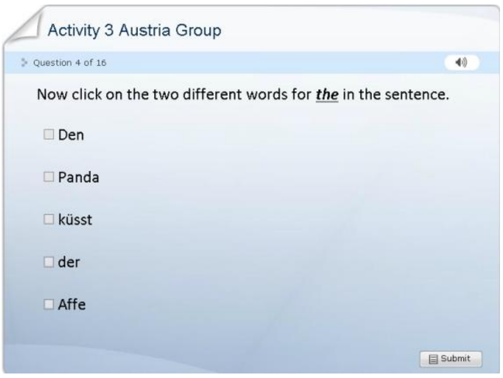
Noticing task:



Transcript: Der Vater ruft den Jungen. [the-NOM father calls the-ACC boy]

Enriched input task (OVS):

Noticing task:



Transcript: Den Panda küsst der Affe [the-ACC panda kisses the-NOM monkey]

Worksheet version:

SVO sentence:

Der Farmer füttert den Hahn. Sensible Silly
[The-NOM farmer feeds the-ACC hen]

OVS sentence:

Den Mann verfolgt der Computer. Sensible Silly
[The-ACC man chases the-NOM computer]

Feedback for worksheet version:

SVO sentence:

Der Farmer füttert den Hahn. Sensible Silly

This sounds ***sensible.*** It is normal for the farmer to feed the hen.

OVS sentence:

Den Mann verfolgt der Computer. Sensible Silly

This sounds ***silly***. It is not normal for a computer to chase a man because computers can’t run!

Appendix 9 Control group vocabulary practice

Missing Words

Monika has dropped her work in the mud! Now Monika can't see some of the words in these German sentences.

Help Monika decide which noun fits in each gap, so that the sentence matches the picture. Write your answer in the gap.

1.	Der Farmer milkt die ______________ .	Hahn	Kuh	
2.	Der ______________ umarmt das Schaf.	Junge	Mädchen	
3.	Der Fisch küsst das ______________ .	Känguru	Kaninchen	
4.	Der ______________ füttert das Pferd.	Tante	Onkel	
5.	Der Mann umarmt das ______________ .	Hahn	Schwein	

...

Sensible or Silly?

Marie is writing sentences in German. But her naughty brother has decided to write some of his own *silly* sentences and mix them up with Marie's ***sensible*** sentences!

Read each sentence and decide whether it is a *sensible* sentence or a *silly* sentence.

1. **Der Opa umarmt das Kind.** Sensible Silly

2. **Der Bär begrüßt die Frau.** Sensible Silly

...

Now you're going to listen to 6 sentences. Listen carefully and decide whether the sentences are *sensible* or *silly*.

7. Sensible Silly

8. Sensible Silly

...

Transcript:

[Der Hamster verfolgt die Katze.]

[Der Onkel begrüßt die Tante.]

Appendix 10 Vocabulary used in the oral tasks

Nouns (animals)	Nouns (items)	Verbs
Affe	Ball	kickt
Bär	Stuhl	küsst
Elefant	Tisch	putzt
Hund		schlägt / trifft
Löwe		umarmt
Tiger		verfolgt

Appendix 11 Number of OVS sentences produced on the Act-Out Production task

Group	Post-test		Delayed post-test	
	N[a]	**OVS**	***N***[a]	**OVS**
TE-FMC	9	35	11	40
TE-F	4	9	8	17

[a]Number of participants utilising OVS word order

Appendix 12 Coding scheme for Sentence Reconstruction task

Subject	*refer to article as being used with the subject*
Object	*refer to article being used with the object*
Doing	*refer to article being used with noun that is 'doing' the action*
Receiving	*refer to article being used with noun that is 'receiving' the action*
One	*refer to feminine or neuter only having one word for the subject and object*
Opposite	e.g. *have **der** for subject so **die** or **das** would be object (or vice versa)*
OVS	*word order can be reversed without changing the meaning.*
Sound	*sounds (or looks) right in a particular order*
Masculine	*refer to article as being used with masculine nouns*
Feminine	*refer to article as being used with feminine nouns*
Neuter	*refer to article being used with neuter nouns*
Male / Boy	*refer to particular article as being used with male words / for boys*
Female / Girl	*refer to particular articles as being used with female words / for girls*
Boy/Girl	***das** can be used with a boy or a girl*
Person	*refer to article as being used with a person*
Object(thing)	*refer to article as being used with an object (i.e. thing)*
den(a)	***den** is the German for **a***
FemE	*using rule that most feminine nouns end in -e*
WO	*refer to SVO word order (e.g. dog is chasing so has to go at beginning)*
MetaL	*use of metalanguage in explanation*
Elim	*use process of elimination (e.g. knew **der** went with XXX, so **die** must go with XXX)*
Prompt	*able to give explanation after prompting from researcher*
EngT	*use English translation to work out the order of the words*

Appendix 13 Instructions for second marker (Sentence Reconstruction task)

INSTRUCTIONS: In this task the pupils are given five words and a picture and are asked to put the words in order, to make a sentence to describe the picture. The pupils are then asked questions about why they chose that particular order. The task is aimed at finding out what the pupils know about the German masculine definite articles ***der*** and ***den***. Points are awarded for the explanations the pupils give for these articles and the position they have put them in in the sentence. The explanations of interest relate particularly to the pupils' understanding that the masculine nominative article ***der*** is used with the Subject of the sentence, and that the masculine accusative article ***den*** is used with the Object of the sentence. Although some pupils might also give explanations about the gender of the nouns (e.g. 'Hund is a masculine noun so I used ***der***'), no points are awarded for explanations of this nature as the focus is on the Subject/Object difference between ***der*** and ***den***. There are three sentences in total. The first sentence uses two masculine nouns plus both articles ***der*** and ***den***. Sentence 2 uses one masculine and one feminine noun and the masculine nominative article ***der*** plus ***die*** (the feminine definite article). Sentence 3 uses one masculine and one neuter noun and the masculine accusative article ***den*** plus ***das*** (the neuter definite article). One point is available for the explanation given for each article in the sentence (2 points in total per sentence). The scoring system for each sentence is as follows:

Sentence 1 (Correct Order: *der Vogel verfolgt den Hund* OR *den Hund verfolgt der Vogel*)

1 point for correctly explaining that ***der*** is used with the SUBJECT of the sentence (the thing DOING the action).

1 point for correctly explaining that ***den*** is used with the OBJECT of the sentence (the thing RECEIVING the action).

½ point given if prompting needed to reach correct explanation for one of the articles.

0 points given if not able to give correct explanation for either article.

For sentences 2 and 3 it should be noted that the feminine article ***die*** (Sentence 2) and the neuter article ***das*** (Sentence 3) can be used with both the SUBJECT and the OBJECT. In order to receive the point for their explanation of these articles, the pupils need to explicitly say whether *in that particular sentence* ***die*** or ***das*** is being used for the SUBJECT of the OBJECT.

Sentence 2 (Correct Order: *der Clown erschreckt die Frau* OR *die Frau erschreckt der Clown*)

1 point for correctly explaining that ***der*** is used with the SUBJECT of the sentence (the thing DOING the action).

1 point for correctly explaining that *in this sentence* ***die*** is used with the OBJECT of the sentence (the thing RECEIVING the action) (because we already know the ***der*** is with the SUBJECT.)

½ point given if prompting needed to reach correct explanation for one of the articles.

0 points given if not able to give correct explanation for either article.

Sentence 3 (Correct Order: *das Kind umarmt den Teddybären* OR *den Teddybären umarmt das Kind*)

1 point for correctly explaining that ***den*** is used with the OBJECT of the sentence (the thing RECEIVING the action).

1 point for correctly explaining that *in this sentence* ***das*** is used with the SUBJECT of the sentence (the thing DOING the action) (because we already know the ***den*** is with the OBJECT.)

½ point given if prompting needed to reach correct explanation for one of the articles.

0 points given if not able to give correct explanation for either article.

Appendix 14 Protocol for oral tasks (set A)

<u>Before starting activities</u>

Make sure all recording equipment is set up and positioned correctly.

- Video camera = need to see soft toys and the researcher's / pupil's hands (NB. In Activities 1 to 3, make sure that when acting out sentences, the toys in use can be seen on video camera).
- Dictaphone= positioned just to one side between researcher and pupil so that both can be heard clearly.
- Laptop = make sure that first sound file is loaded and second is easily accessible.

<u>Make sure that pupils feel comfortable:</u>

R: Today we're going to do four short activities that will help us practise listening to and saying German sentences, and we will be using the soft toys to help us understand the sentences. I am going to record the activities on the video camera and the dictaphone, so that later I can remember what we did, otherwise I will forget! Is that ok?

R: Let's just check that the dictaphone is working. When I say 'Go', I want you to say 'Hello. My name is…'. *(Press record button and record pupil introducing themselves. Play back to pupil afterwards.)*

R: Ok great. Now we are going to quickly have a look at the words we will need when we're doing the activities. I'm going to show you a list of the nouns that we will use, and I want you to have a look at them and check that you know them all.

(Give list of nouns to pupils – allow up to 30 seconds to look through)

R: Ok, let's just check we know them. I'm going to say the name of each animal or thing and I want you to point to which one it is. *(Say each name in turn and after each one ask pupil to point to the right animal / thing).* Ok, now I am going to point to one of the animals or things and I want you to tell me the German name for it. *(Point to each animal / thing in turn and after each one ask pupil to tell you the German name for it).*

R: *(Same procedure for verbs)* Now we are going to look at the verbs we are going to use today. Here is the list of verbs that we will use, and I want you to have a look at them and check that you know them all. *(Give pupils up to 30 seconds to look through the list)*

R: Now I'm going to show you the action we will use for each verb. *(Using the bear and the elephant act out each of the 6 verbs in turn, and tell the pupil which verb it is for).*

R: Now let's just check that we know them all. I'm going to say a verb and I want you to show me the action for it. *(Say each verb in turn and after each one ask pupil to act out the verb using the bear and elephant).*Great, now I'm going to act out the verbs and I want you to tell me which verb it is. *(Act out each verb in turn using the bear and elephant and after each one ask pupil to tell you the German word for that verb. NB: If pupil unsure of correct verb or chooses wrong verb, then prompt (are you sure?) / correct them. Pupils can look at list if needed).*

Ok, let's get started. Activity 1 …

NB: Switch off video camera and Dictaphone at end of Activity 4.

<u>Activity 1: Sentence repetition task</u> **(Sound file: Sentence Rep 1)**

In this activity, the pupils will watch an action, acted out with the soft toys and then listen to a sentence which describes the action. The sentence will be followed by a beep; after the beep the pupils must repeat the sentence.

Instructions for researcher:

1) Say ‘Activity 1’ aloud.
2) Read instructions to pupil.
3) Say ‘Number …’ aloud before each new action/sentence.
4) Act out the sentence using the soft toys.
5) Play sentence, pause after ‘beep’.
6) After pupil has repeated the sentence, go back to 3) for next sentence.

Instructions for pupil:

In this activity you are going to be listening to and repeating some German sentences. This activity is going to help us to practise saying German words and sentences.

I am going to show you an action using the toys and then we are going to listen to a German sentence. After the sentence there will be a beep. After the beep you just have to repeat the sentence that you heard.

Let’s practise using an English sentence.

R: *Act-out;* The bear chases the elephant.

R: *Say;* The bear chases the elephant (beep).

P: *Repeat sentence.*

Ok, let’s start the activity.

Sentence 1: *Act-out; the dog kisses the elephant.* (Play sentence 1, pause after beep)

Sentence 2: *Act-out; the bear chases the lion.* (Play sentence 2, pause after beep)

…

<u>Activity 2 (Act-out Comprehension task)</u> **(Sound file: Act-out Comp 1)**

In this activity, the pupils will listen to a sentence and then act-out the sentence using the soft toys.

Instructions for researcher:

1) Say ‘Activity 2’ aloud at the start.
2) Read instructions to pupil.
3) Say ‘Number …’ aloud before each new action / sentence.
4) Play sentence to pupil, pause after each sentence.
5) After pupil has acted out the sentence using the toys, go back to 3) for next sentence.

Instructions for pupil:

In this activity we are going to practise listening to German sentences. I am going to play you a sentence, and I would like you to act out the sentence using the soft toys. Make sure you listen to what the sentence is telling you; some of the actions might seen a bit strange, but that's ok! Let's practise using an English sentence.

R: *Say;* The table chases the elephant.

P: *Act-out sentence using soft toys.*

Ok, let's start the activity.

(Correct actions)

1) *The monkey chases the lion.*
2) *The dog hugs the tiger.*

…

<u>Activity 3 (Act-out Production task)</u> **(Sound file: None)**

In this activity, the pupils will have to create a sentence to describe an action.

Instructions for researcher:

1) Say 'Activity 3' aloud at the start.
2) Read instructions to pupil.
3) Say 'Number …' aloud before each new action / sentence.
4) Act-out sentence.
5) After pupil has produced a sentence, go back to 3) for next sentence.

Instructions for pupil:

In this activity we are going to practise making German sentences to describe different actions. I am going to show you an action using the soft toys, and I would like you to make a sentence to describe the action; just like the sentences we have been using in the other activities. Let's do a practice sentence first.

R: *Act;* *The bear hugs the elephant.*

R: *Say;* The sentence for this action could be, 'Der Bär umarmt den Elefanten'.

Ok, let's start the activity.

(NB: You can give the pupil a noun or a verb if they get stuck, but not the whole sentence. Allow ~ 5 seconds before asking if they need help with a word)

Actions:

1) *The bear kisses the monkey.*
2) *The lion chases the ball.*

…

Activity 4 (Sentence Reconstruction task) **(Sound file: None)**

In this activity, the pupils will be shown a picture and five words. Their job is to put the words into the correct order so that they make a sentence to describe the picture.

The aim of the activity is to find out **WHY** the pupils put the words into that order (particularly the words 'der' and 'den' (two words for 'the' in German).

NB: There are no full stops or capital letters to help the pupils work out the answer; with the exception of the two nouns, as in German all nouns must start with a capital letter.

Instructions for researcher:

1) Say 'Activity 4' aloud at the start.
2) Read instructions to pupil.
3) Say 'Number …' aloud before each new action / sentence.
4) Show picture and words to pupils.
5) While / After pupil has arranged words into a sentence, discuss why they chose that order.

Instructions for pupil:

In this activity we are going to practise making German sentences to describe some pictures. I am going to show you a picture and five words. Your job is to put the words into order so that they make a sentence to describe the picture. Ok, let's start the activity.

Order of pictures:

1) The dog chasing a cat.	Correct order: **der** Hund verfolgt die Katze
2) The man writing a letter.	Correct order: **der** Mann schreibt **den** Brief
3) The frisbee hitting a boy.	Correct order: das Frisbee trifft **den** Jungen.

NB: Prompt the pupil to find out why they have chosen the order. If the pupil chooses the wrong order, don't tell them that it is wrong; discuss **WHY** they chose that order. (They may correct on their own, if they don't then that is ok. The important thing is to find out **WHY** they chose that order.)

Questions: Why have you chosen that order for the words?
Why have you put '**der**' in that position?
Why have you put '**den**' in that position?
How did you know that that word goes with **der**/**den**/**die**/**das**?

Ask pupils to explain points a bit more, for example if they say 'because of the gender' or if they say 'because it is a thing/object' ask, 'can you tell me a bit more about the gender / that?'.

Appendix 15 Results of the Shapiro-Wilk test of normality

By experimental group

	Task	Time	Group	W	df	*p*
Written	Sentence Matching	Pre	TE-FMC	0.907	45	*.002***
			TE-F	0.891	38	*.001***
			Control	0.943	50	*.018**
		Post	TE-FMC	0.822	45	*.000***
			TE-F	0.833	38	*.000***
			Control	0.628	50	*.000***
		Delayed Post	TE-FMC	0.838	45	*.000***
			TE-F	0.803	38	*.000***
	Gap-Fill	Pre	TE-FMC	0.626	45	*.000***
			TE-F	0.574	38	*.000***
			Control	0.565	50	*.000***
		Post	TE-FMC	0.886	45	*.000***
			TE-F	0.832	38	*.000***
			Control	0.514	50	*.000***
		Delayed Post	TE-FMC	0.810	45	*.000***
			TE-F	0.866	38	*.000***
Oral	Act-Out Comprehension	Pre	TE-FMC	0.768	45	*.000***
			TE-F	0.851	41	*.000***
			Control	0.845	46	*.000***
		Post	TE-FMC	0.868	45	*.000***
			TE-F	0.870	41	*.000***
			Control	0.571	46	*.000***
		Delayed Post	TE-FMC	0.858	45	*.000***
			TE-F	0.845	41	*.000***
	Act-Out Production	Pre	TE-FMC	0.684	45	*.000***
			TE-F	0.629	41	*.000***
			Control	0.688	46	*.000***
		Post	TE-FMC	0.781	45	*.000***
			TE-F	0.899	41	*.002***
			Control	0.669	46	*.000***
		Delayed Post	TE-FMC	0.828	45	*.000***
			TE-F	0.857	41	*.000***
	Sentence Repetition	Pre	TE-FMC	0.944	45	*.031**
			TE-F	0.920	41	*.007***
			Control	0.910	46	*.002***
		Post	TE-FMC	0.930	45	*.009***
			TE-F	0.904	41	*.002***
			Control	0.911	46	*.002***
		Delayed Post	TE-FMC	0.887	45	*0.000***
			TE-F	0.930	41	*0.015**

*significant at the .05 level
**significant at the .01 level

By experimental group *(continued)*

Task		Time	Group	W	df	*p*
Metalinguistic	Sentence Reconstruction (Order)	Pre	TE-FMC	0.829	45	*.000***
			TE-F	0.821	41	*.000***
			Control	0.833	46	*.000***
		Post	TE-FMC	0.523	45	*.000***
			TE-F	0.442	41	*.000***
			Control	0.808	46	*.000***
		Delayed Post	TE-FMC	0.613	45	*.000***
			TE-F	0.618	41	*.000***
	Sentence Reconstruction (Explanation)	Pre	TE-FMC	0.209	45	*.000***
			TE-F	-	-	-
			Control	-	-	-
		Post	TE-FMC	0.788	45	*.000***
			TE-F	0.767	41	*.000***
			Control	0.209	46	*.000***
		Delayed Post	TE-FMC	0.857	45	*.000***
			TE-F	0.932	41	*.017**

*significant at the 0.05 level
**significant at the 0.01 level

By age group *(pre-test only)*

Task		Group	W	df	*p*
Written	Sentence Matching	Year 6	0.930	66	*.001***
		Year 5	0.900	67	*.000***
	Gap-Fill	Year 6	0.564	66	*.000***
		Year 5	0.560	67	*.000***
Oral	Act-Out Comprehension	Year 6	0.743	66	*.000***
		Year 5	0.890	66	*.000***
	Act-Out Production	Year 6	0.644	66	*.000***
		Year 5	0.698	66	*.000***
	Sentence Repetition	Year 6	0.939	66	*.003***
		Year 5	0.936	66	*.002***
Metalinguistic	Sentence Reconstruction (O)[86]	Year 6	0.861	66	*.000***
		Year 5	0.809	66	*.000***
	Sentence Reconstruction (E)[87]	Year 6	0.103	66	*.000***
		Year 5	0.103	66	*.000**

*significant at the 0.05 level
**significant at the 0.01 level

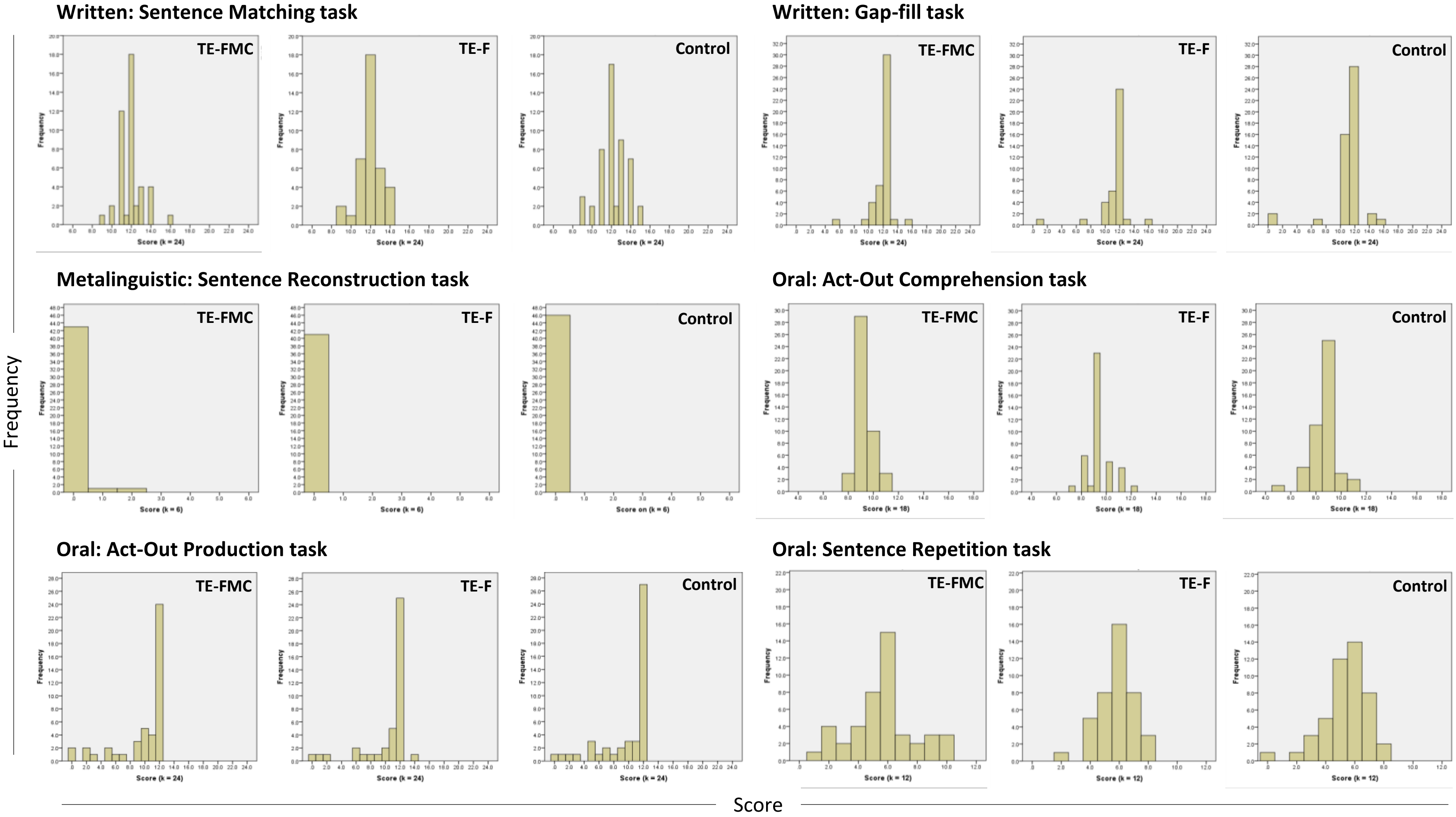
[86] Sentence Reconstruction (O) refers to the score given for correctly ordering of the words in the sentence.
[87] Sentence Reconstruction (E) refers to the score given for the explanation regarding the position of the articles in the sentence.

By school *(pre-test only)*

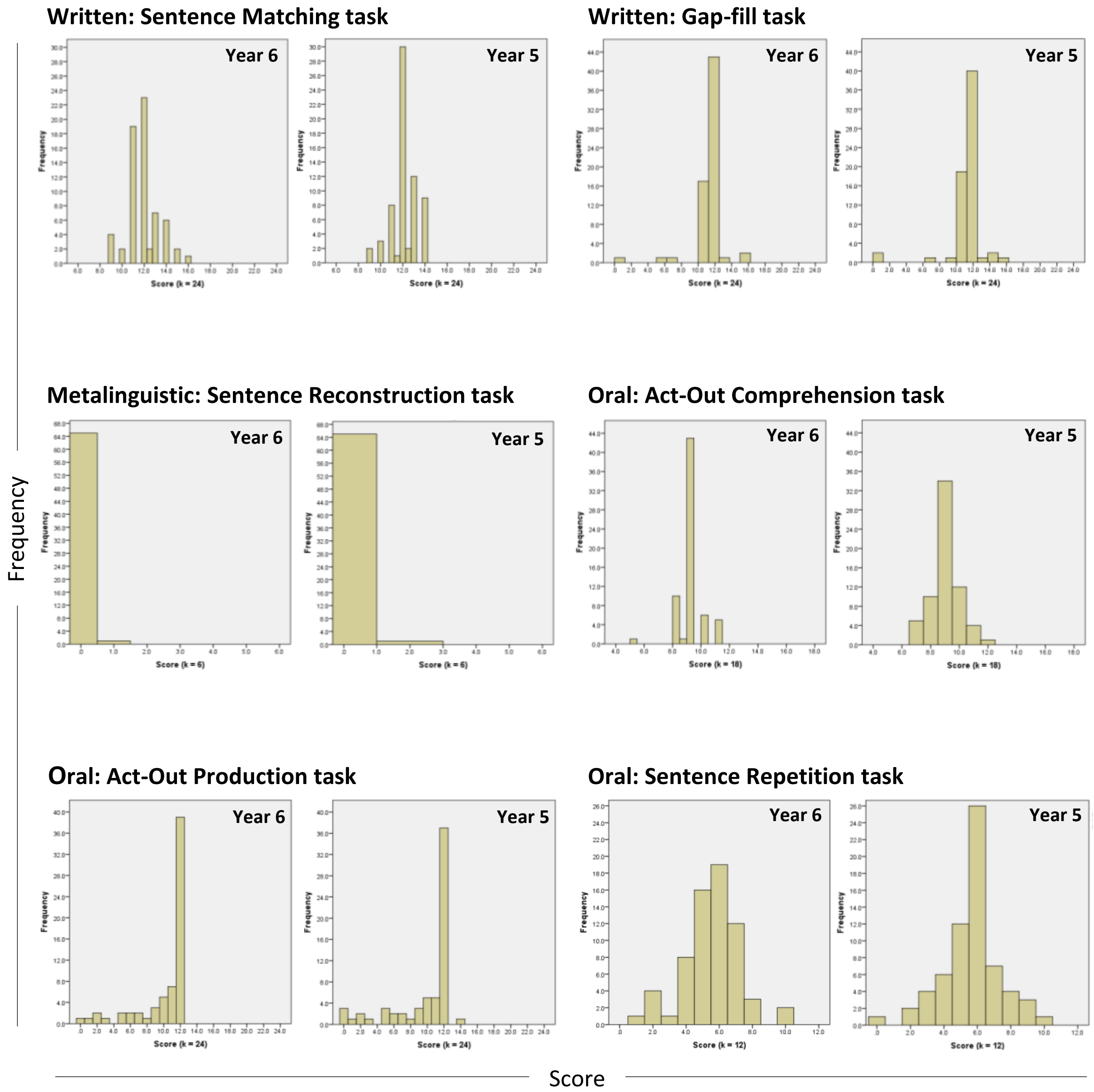
	Task	School	W	df	*p*
Written	**Sentence Matching**	1	0.904	36	*.004***
		2	0.922	76	*.001***
		3	0.94	25	*0.151*
	Gap-fill	1	0.609	36	*.001***
		2	0.584	72	*.001***
		3	0.511	25	*.001***
Oral	**Act-Out Comprehension**	1	0.802	38	*.001***
		2	0.848	72	*.001***
		3	0.726	22	*.001***
	Act-Out Production	1	0.682	38	*.001***
		2	0.703	72	*.001***
		3	0.583	22	*.001***
	Sentence Repetition	1	0.947	38	*0.072*
		2	0.941	72	*.002***
		3	0.923	22	*0.089*
Metalinguistic	**Sentence Reconstruction (O)**	1	0.783	38	*.001***
		2	0.867	72	*.001***
		3	0.842	22	*.002***
	Sentence Reconstruction (E)	1	0.235	38	*.001***
		2	-	-	-
		3	-	-	-

*significant at the 0.05 level
**significant at the 0.01 level

Appendix 16 Distribution of scores on each task at pre-test (by experimental group)

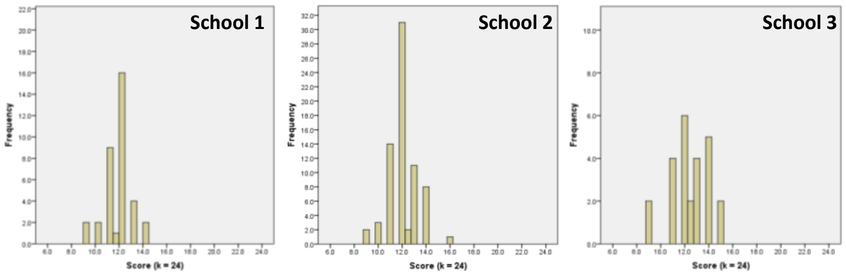


Appendix 17 Distribution of scores on each task at pre-test (by age group)

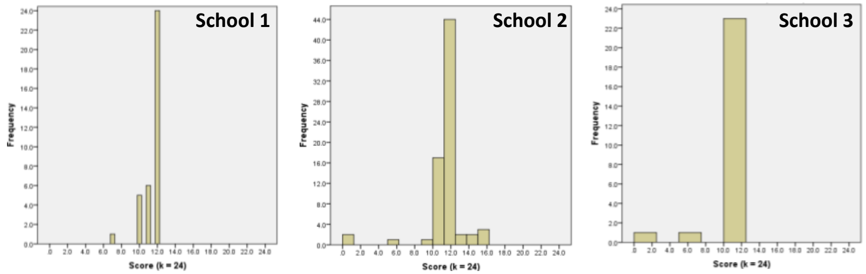


Appendix 18 Distribution of scores on each task at pre-test (by school)

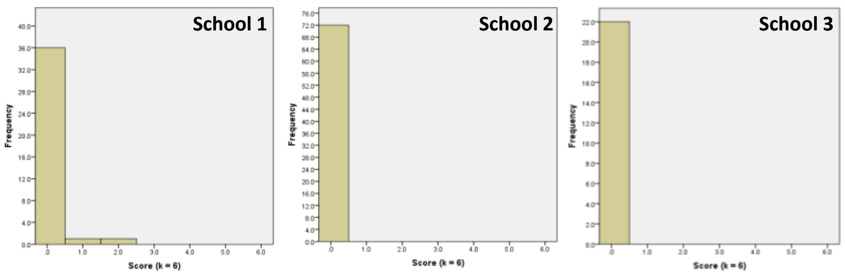
Written: Sentence Matching task



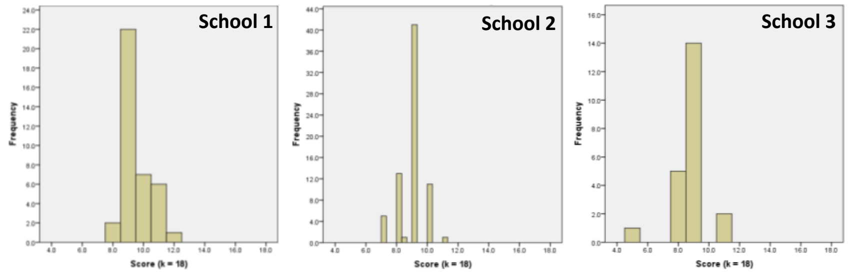
Written: Gap-fill task



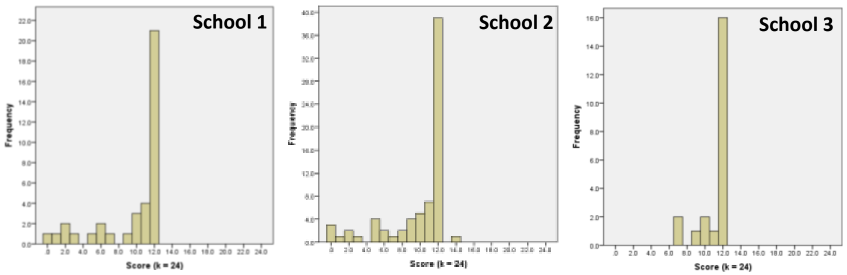
Metalinguistic: Sentence Reconstruction task



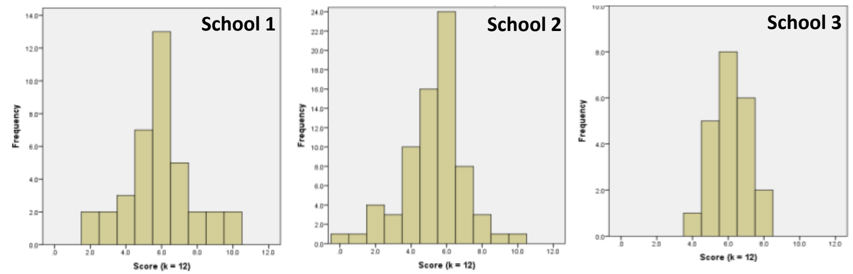
Oral: Act-Out Comprehension task



Oral: Act-Out Production task



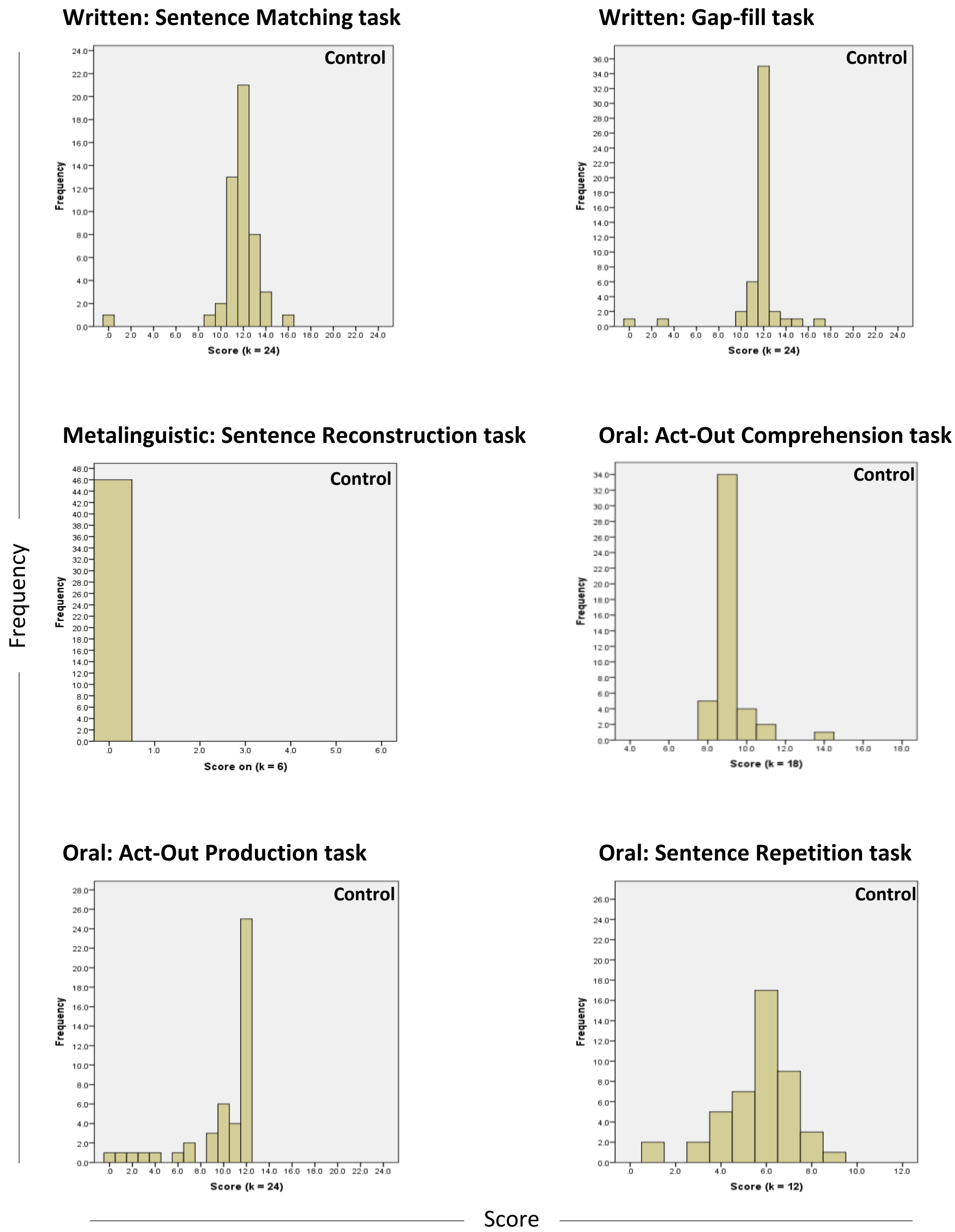
Oral: Sentence Repetition task



Frequency

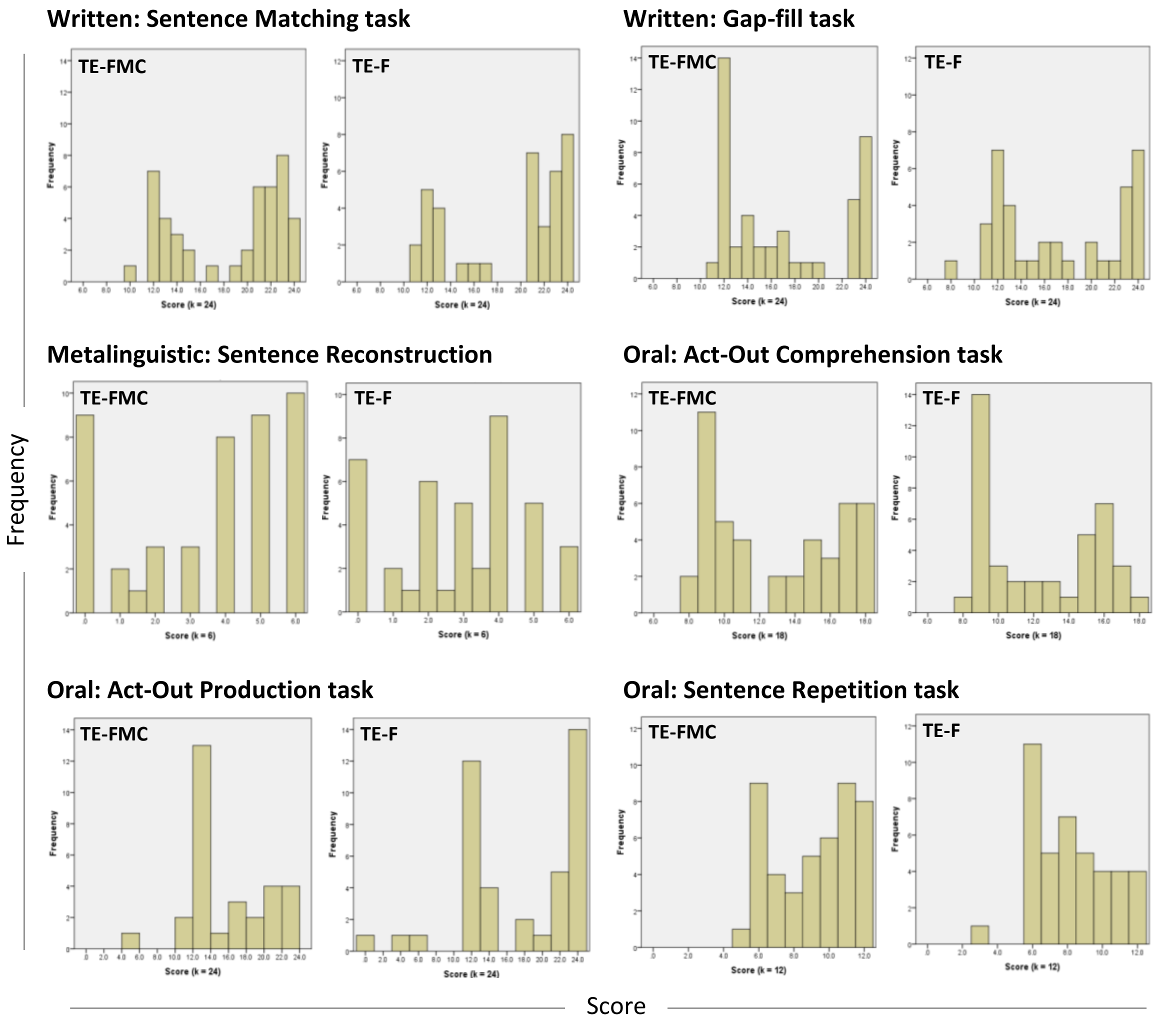
Score

Appendix 19 Distribution of scores on each task at post-test (Control group only[88])



[88] See Figures 5.1a, 5.1b, 5.1c for histograms of TE-FMC and TE-F groups' post-test scores at post-test

Appendix 20 Distribution of scores on each task at delayed post-test



Appendix 21 Results of Levene's test of homogeneity of variance

By experimental group

Task		Time	F	*p*
Written	**Sentence Matching**	**Pre**	1.376	*0.256*
		Post	54.788	*0.000***
		Delayed post	0.113	*0.738*
	Gap-Fill	**Pre**	1.055	*0.351*
		Post	55.607	*0.000***
		Delayed post	0.426	*0.516*
Oral	**Act-Out Comprehension**	**Pre**	1.418	*0.246*
		Post	57.505	*0.000***
		Delayed post	1.517	*0.222*
	Act-Out Production	**Pre**	0.598	*0.552*
		Post	25.04	*0.000***
		Delayed post	1.502	*0.224*
	Sentence Repetition	**Pre**	4.202	*0.017**
		Post	3.526	*0.032**
		Delayed post	0.539	*0.465*
Metalinguistic	**Sentence Reconstruction (O)**	**Pre**	0.187	*0.83*
		Post	5.844	*0.004***
		Delayed post	0.016	*0.900*
	Sentence Reconstruction (E)	**Pre**	7.631	*0.001***
		Post	39.954	*0.000***
		Delayed post	3.462	*0.066*

*significant at the .05 level
**significant at the .01 level

By age group *(pre-test only)*

Task		F	*p*
Written	**Sentence Matching**	1.501	*.223*
	Gap-Fill	0.104	*.748*
Oral	**Act-Out Comprehension**	2.244	*.137*
	Act-Out Production	1.881	*.173*
	Sentence Repetition	0.011	*.916*
Metalinguistic	**Sentence Reconstruction (O)**	1.889	*.172*
	Sentence Reconstruction (E)	0.825	*.365*

By school *(pre-test only)*

Task		F	*p*
Written	**Sentence Matching**	2.222	*.112*
	Gap-Fill	1.107	*.334*
Oral	**Act-Out Comprehension**	1.325	*.27*
	Act-Out Production	5.474	*.005***
	Sentence Repetition	1.902	*.153*
Metalinguistic	**Sentence Reconstruction (O)**	0.794	*.454*
	Sentence Reconstruction (E)	-	-

Appendix 22 Participant performance by age group

Descriptive statistics *(pre-test only)*

Age group	Sentence Matching (*k* = 24)				Gap-Fill (*k* = 24)			
	N	***M***	***SD***	***Mdn***	***N***	***M***	***SD***	***Mdn***
Year 6	66	11.93	1.41	12	66	11.41	1.99	12
Year 5	67	12.16	1.14	12	67	11.29	2.23	12

Age group	Act-Out Competition (*k* = 18)				Act-Out Production (*k* = 24)			
	N	***M***	***SD***	***Mdn***	***N***	***M***	***SD***	***Mdn***
Year 6	66	9.02	0.90	9	66	10.21	3.10	12
Year 5	66	9.05	1.01	9	66	9.80	3.63	12

Age group	Sentence Repetition (*k* = 12)				Sentence Reconstruction (*k* = 6)			
	N	***M***	***SD***	***Mdn***	***N***	***M***	***SD***	***Mdn***
Year 6	66	5.55	1.70	6	66	4.26	1.45	4
Year 5	66	5.67	1.75	6	66	4.11	1.28	4

Age group	Sentence Reconstruction (*k* = 6)			
	N	***M***	***SD***	***Mdn***
Year 6	66	0.02	0.12	0
Year 5	66	0.03	0.25	0

Results of Mann Whitney U test with between-group variable Age *(pre-test only)*

Task	*U*	*z*	*p*	*r*
Sentence Matching	2530.00	1.485	.137	.13
Gap-Fill	2104.00	-0.546	.585	-.05
Act-Out Comprehension	2220.00	0.214	.830	.02
Act-Out Production	2111.00	-0.339	.734	-.03
Sentence Repetition	2271.00	0.044	.664	.04
Sentence Reconstruction (O)	1974.50	-1.001	.317	.09
Sentence Reconstruction (E)	2178.50	0.011	.991	.00

Appendix 23 Participant performance by school

Descriptive statistics *(pre-test only)*

School	Sentence Matching (k = 24)				Gap-Fill (k = 24)			
	N	*M*	*SD*	*Mdn*	*N*	*M*	*SD*	*Mdn*
1	36	11.68	1.12	12	36	11.42	1.05	12
2	72	12.08	1.21	12	72	11.48	2.32	12
3	25	12.46	1.58	12.7	25	10.87	2.54	12

School	Act-Out Competition (k = 18)				Act-Out Production (k = 24)			
	N	*M*	*SD*	*Mdn*	*N*	*M*	*SD*	*Mdn*
1	38	9.53	0.92	9	38	9.66	3.71	12
2	72	8.86	0.81	9	72	9.83	3.55	12
3	22	8.77	1.15	9	22	11.18	1.59	12

School	Sentence Repetition (k = 12)				Sentence Reconstruction (O) (k = 6)			
	N	*M*	*SD*	*Mdn*	*N*	*M*	*SD*	*Mdn*
1	38	5.90	1.89	6	38	4.32	1.16	4
2	72	5.29	1.75	6	72	4.08	1.44	4
3	22	6.14	1.04	6	22	4.27	1.45	4

School	Sentence Reconstruction (E) (k = 6)			
	N	*M*	*SD*	*Mdn*
1	38	0.08	0.36	0
2	72	0	0	0
3	22	0	0	0

Results of Kruskall Wallis test with between-group variable School *(pre-test only)*

Task	*H*	*df*	*p*
Sentence Matching	6.560	2	.037[89]
Gap-Fill	1.902	2	.386
Act-Out Comprehension	13.405	2	.001[90]
Act-Out Production	2.612	2	.271
Sentence Repetition	5.835	2	.054
Sentence Reconstruction (O)	0.543	2	.762
Sentence Reconstruction (E)	4.985	2	.083

[89] School 1 *vs.* School 3; $z = -2.556$, p .011, $r = -.36$

[90] School 1 *vs.* School 2; $z = 3.391$, $p = .001$, $r = .32$. School 1 *vs.* School 3, $z = 2.864$, $p = .004$, $r = .37$

Appendix 24 Descriptive statistics for Animacy conditions

Written: Sentence Matching task

Group	N	A+A (*k* = 8)						A+I (*k* = 8)						I+A (*k* = 8)					
		Pre-test			Post-test			Pre-test			Post-test			Pre-test			Post-test		
		M	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	45	4.03	0.66	4	6.64	1.90	8	3.95	0.60	4	6.44	1.87	7	3.84	0.95	4	6.22	1.69	6
TE-F	38	4.05	0.61	4	6.61	1.64	7	4.03	0.54	4	6.79	1.54	7	3.89	0.83	4	6.26	1.72	7
Control	50	4.18	0.65	4	4.06	0.62	4	4.00	0.88	4	3.82	0.60	4	4.00	0.83	4	4.08	0.75	4
Total	133	4.09	0.64	4	5.66	1.91	5	3.99	0.70	4	5.56	1.95	5	3.92	0.87	4	5.43	1.76	5

Written: Gap-fill task

Group	N	A+A (*k* = 12)						A+I (*k* = 12)					
		Pre-test			Post-test			Pre-test			Post-test		
		M	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	45	5.81	0.68	6	6.60	2.93	6	5.74	0.93	6	6.62	2.80	6
TE-F	38	5.66	1.19	6	8.97	2.98	10	5.68	1.14	6	9.26	2.71	10
Control	50	5.47	1.27	6	5.80	1.11	6	5.66	1.55	6	5.82	1.42	6
Total	133	5.64	1.09	6	6.98	2.74	6	5.69	1.25	6	7.08	2.73	6

Oral: Act-Out Production task

Group	N	A+A (*k* = 6)						A+I (*k* = 6)					
		Pre-test			Post-test			Pre-test			Post-test		
		M	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*	*M*	SD	*Mdn*
TE-FMC	45	0.00	0.00	0	3.16	2.77	4	0.00	0.00	0	3.04	2.69	4
TE-F	41	0.20	0.00	0	2.59	2.67	1	0.41	1.18	0	2.73	2.63	2
Control	46	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0	0.00	0.00	0
Total	132	0.06	0.37	0	1.88	2.59	0	0.13	0.68	0	1.89	2.54	0

Appendix 25 Post-test scores for Not Got It sub-group on Sentence Reconstruction (Explanation) task

Participant	Group	S. Recon (E) (k = 6)	S. Match (k = 24)	Gap-fill (k = 24)	AO Comp (k = 18)	AO Prod (k = 24)	S. Rep (k = 12)
10	**TE-FMC**	**1**	14	21+	15	24+	9
12	**TE-FMC**	**2**	22+	16	10	12	10
47	**TE-FMC**	**1**	24+	24+	9	12	11
65	**TE-FMC**	**2**	10	8	10	11	8
26	**TE-F**	**2**	-	-	6	11	12+
41	**TE-F**	**0**	13	12	9	9	7
42	**TE-F**	**0**	17	12	11	12	5
43	**TE-F**	**0**	17	24+	16	24+	9
44	**TE-F**	**0**	-	-	9	0	0
78	**TE-F**	**0**	21+	21+	11	12	8
80	**TE-F**	**2**	16	13	9	12	8
85	**TE-F**	**1**	10	11	9	12	9

+Ceiling-level (i.e. Got It) performance

Appendix 26 Descriptive statistics for TE-FMC intervention (week 1)

Activity	Number of items	Score		
		Mean	SD	Median
Who's doing what?	**12**	10.68	1.94	12
Missing Nouns	**20**	16.30	3.24	16.5
Which picture?	**16**	14.08	2.36	15

N = 40

Definitions

A	Animate noun
A+A	Sentence containing two animate nouns
A+I	Sentence containing animate subject and inanimate object
ACC	Accusative case
ANOVA	Analysis of variance
ANCOVA	Analysis of covariance
AoA	Age of acquisition
CI	Confidence interval
CPH	Critical Period Hypothesis
DfE	Department for Education
DfEE	Department for Education and Employment
DfES	Department for Education and Skills
EI	Enriched Input
FDH	Fundamental Difference Hypothesis
FMC	Form-meaning connection
FNP	First Noun Principle
GJT	Grammaticality judgement task
I+A	Sentence containing inanimate subject and animate object
k	Number of test items
K	Total number of test items
KMO	Kaiser-Meyer-Olkin statistic
KS1	Key Stage 1
KS2	Key Stage 2
KS3	Key Stage 3
KS4	Key Stage 4
L1	First / native language
L2	Second / foreign language
m+f/n	Sentence containing one masculine and one feminine or neuter noun
m+m	Sentence containing two masculine nouns

M	Mean
Mdn	Median
MLAT-E	Modern Language Aptitude Test - Elementary
MOI	Meaning-based output instruction
n	Number of participants
N	Total number of participants
NOM	Nominative case
NVN	Noun-verb-noun
OVS	Object-verb-subject
OVS+Obj	Item in object-verb-subject word order targetting the object
OVS+Subj	Item in object-verb-subject word order targetting the subject
PCA	Principal Component Analysis
PI	Processing Instruction
QCA	Qualifications and Curriculum Authority
r	Pearson's correlation co-effiecient
SATs	Standard Assessment Tests
SD	Standard deviation
SI	Structured input
SLA	Second language acquisition
SVO	Subject-verb-object
SVO+Obj	Item in subject-verb-object word order targetting the object
SVO+Subj	Item in subject-verb-object word order targetting the subject
t	Test score
TE-F	Task-essential form only (intervention)
TE-FMC	Task-essential form-meaning connection (intervention)
UG	Universal grammar
WO	Word order

References

Agiasophiti, Z. (2013). Exploring possible effects of gender and enhanced versus unenhanced processing instruction on the acquisition of case marking in L2 German. In J. F. Lee & A. Benati (Eds.), *Individual differences and processing instruction* (pp. 153-184). Sheffield: Equinox Publishing.

ALA. (2012). Association for language awareness: About. Retrieved 10/02/2015, from http://www.languageawareness.org/web.ala/web/about/tout.php

Alderson, J. C., Clapham, C., & Steel, D. (1997). Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching, 1*(2), 93-121.

Allen, L. Q. (2000). Form-meaning connections and the French causative: An experiment in processing instruction. *Studies in Second Language Acquisition, 22*(1), 69-84.

Anderson, J. R. (1992). Automaticity and the ACT theory. *American Journal of Psychology, 105*(2), 164-180.

Anderson, J. R. (2005). *Cognitive psychology and its implications*. New York: Worth Publishers.

Andrews, R., Torgerson, C., Beverton, S., Freeman, A., Locke, T., Low, G., . . . & Zhu, D. (2004). *The effect of grammar teaching (sentence combining) in English on 5 to 16 year olds' accurcacy and quality in written composition: Review summary.* University of York, UK.

Andrews, R., Torgerson, C., Beverton, S., Locke, T., Low, G., Robinson, A., & Zhu, D. (2004). *The effect of grammar teaching (syntax) in English on 5 to 16 year old's accuracy and quality in written comprehension.* London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Angelovska, T., & Benati, A. (2013). Processing instruction and the age factor: Can adults and school-age native speakers of German process English past tense correctly? In J. F. Lee & A. Benati (Eds.), *Individual differences and processing instruction* (pp. 131-151). Sheffield: Equinox Publishing.

Barton, A., Bragg, J., & Serratrice, L. (2009). 'Discovering language' in primary school: An evaluation of a language awareness programme. *Language Learning Journal, 37*(2), 145-164.

Benati, A. (2001). A comparative study of the effects of processing instruction and output-based instruction on the acquisition of the Italian future tense. *Language Teaching Research, 5*(2), 95-127.

Benati, A. (2004). The effects of processing instruction and its components on the acquisition of gender agreement in Italian. *Language Awareness, 13*(2), 67-80.

Benati, A. (2005). The effects of processing instruction, traditional instruction and meaning-output instruction on the acquisition of the English past simple tense. *Language Teaching Research, 9*(1), 67-93.

Benati, A., & Lee, J. F. (2008). *Grammar acquisition and processing instruction: Secondary and cumulative effects*. Bristol: Multilingual Matters.

Bialystok, E. (1994). Representation and ways of knowing: Three issues in second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 549-569). San Diego: Academic Press.

Birdsong, D. (2005). Understanding age effects in second language acquisition. In J. Kroll & A. de Groot (Eds.), *The handbook of bilingualism: Psycholinguistic approaches* (pp. 109-127). Oxford: Oxford University Press.

Birdsong, D., & Molis, M. (2001). On the evidence of maturational constraints in second language acquisition. *Journal of Memory and Language, 44*(2), 235-249.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis, 20*(1), 3-47.

Bley-Vroman, R. (2009). The evolving context of the fundamental difference hypothesis. *Studies in Second Language Acquisition, 31*(2), 175-198.

Board, K., & Tinsley, T. (2014). *Language trends 2013/2014: The state of language learning in primary and secondary schools in England*. Reading: CfBT Education Trust.

Bouffard, L. A., & Sarkar, M. (2008). Training 8-year-old French immersion students in metalinguistic analysis: An innovation in form-focussed pedagogy. *Language Awareness, 17*(1), 3-24.

Bryant, P., Devine, M., Ledward, A., & Nunes, T. (1997). Spelling with apostrophes and understanding possession. *British Journal of Educational Psychology, 67*(1), 91-110.

Bryant, P., Nunes, T., & Bindman, M. (2000). The relations between children's linguistic awareness and spelling: The case of the apostrophe. *Reading and Writing: An Interdisciplinary Journal, 12*(2), 253-276.

Cable, C., Driscoll, P., Mitchell, R., Sing, S., Cremin, T., Earl, J., . . . & Heins, B. (2012). Language learning at Key Stage 2: Findings from a longitudinal study. *Education 3-13: International Journal of Primary, Elementary and Early Years Education, 40*(4), 363-378.

Cadierno, T. (1995). Formal instruction from a processing perspective: An investigation into the Spanish past tense. *The Modern Language Journal, 79*(2), 179-193.

Cameron, L. (2001). *Teaching languages to young learners*. Cambridge: Cambridge University Press.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Carlisle, J. F. (2000). Awareness of the structure and meaing of morphologically complex words: Impact on reading. *Reading and Writing: An Interdisciplinary Journal, 12*(3), 169-190.

Carroll, J. B. (1971). *Implications of aptitude test research and psycholinguistic theory for foreign language teaching*. Paper presented at the XVIIth International Congress, International Association of Applied Psychology, Liege, Belguim. http://files.eric.ed.gov/fulltext/ED057071.pdf

Carroll, J. B. (1973). Implications of aptitude test research and psycholinguistic theory for foreign language teaching. *International Journal of Psycholinguistics, 2*(1), 5-14.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in langauge learning aptitude* (pp. 83-118). Rowley: Newbury House.

Carroll, J. B., & Sapon, S. (1959). *The modern langauges aptitude test*. San Antonio: Psychological Corporation.

Carroll, J. B., & Sapon, S. (2002). *Modern Language Aptitude Test - Elementary: Manual 2002 edition*. Rockville, MD: Second Language Testing, Inc.

Cenoz, J. (2003). The influence of age on the acquisition of English: General proficiency, attitudes, and code-switching. In M. P. Garcia Mayo & M. L. Garcia Lecumberri (Eds.), *Age and the acquisition of English as a foreign language* (pp. 77-93). Clevedon: Multilingual Matters.

Chan, A., Meints, K., Lieven, E., & Tomasello, M. (2010). Young children's comprehension of English SVO word order revisited: Testing the same children in act-out and intermodal preferential looking tasks. *Cognitive Development, 25*(1), 30-45.

Chaudron, C. (2003). Data collection in SLA research. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 762-828). Malden: Blackwell Publishing.

Cheater, C. (2007). *The Catherine Cheater scheme of work for French (Year 5)*. Wakefield: Teaching and Learning Publications.

Chen, H.-Y. (2009). *A classroom quasi-experimental study to explore processing instruction*. Doctoral dissertation. University of York. York, Uk.

Cheng, A. C. (2004). Processing instruction and Spanish *ser* and *estar*: Forms with semantic-aspectual values. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 119-142). New Jersey: Lawrence Erlbaum Associates, Inc.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

Cohen, L., Manion, L., Morrison, K., & Bell, R. (2011). *Research methods in education* (7th ed.). London: Routledge.

Culman, H., Henry, N., & VanPatten, B. (2009). The role of explicit information in instructed SLA: An online study with processing instruction and German accusative case inflections. *Die Unterrichtspraxis, 42*(1), 19-31.

de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10*(1), 7-21.

de Graaf, R. (1997). *Differential effects of explicit instruction on second language acquisition*. The Hague: Holland Institute of Generative Linguistics.

De Jong, N. (2005). Can second language grammar be learned through listening? An experiment study. *Studies in Second Language Acquisition, 27*(2), 205-234.

Deacon, S. H., & Bryant, P. (2005). What young children do and do not know about the spelling of inflections and derivations. *Developmental Science, 8*(6), 583-594.

Dearing, R., & King, L. (2007). *Languages review*. London: Department for Education and Skills.

DeKeyser, R. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition, 17*(3), 379-410.

DeKeyser, R. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*(2), 195-221.

DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42-63). Cambridge: Cambridge University Press.

DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499-533.

DeKeyser, R. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313-348). Oxford: Blackwell Publishing Ltd.

DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten (Ed.), *Theories in second language acqusiition: An introduction* (pp. 97-113). New Jersey: Lawrence Erlbaum Associates.

DeKeyser, R., & Criado, R. (2012). Automatization, skill acquisition, and practice in second language acquisition. In C. A. Chapelle, (Ed.), *The Encyclopedia of Applied Linguistics*. Oxford: Blackwell Publishing Ltd.

DeKeyser, R., & Larson-Hall, J. (2005). What does the critical period really mean? In J. Kroll & A. de Groot (Eds.), *The handbook of bilingualism: Psycholinguistic approaches* (pp. 88-108). Oxford: Oxford University Press.

DeKeyser, R., Salaberry, R., Robinson, P., & Harrington, M. (2002). What gets processed in processing instruction? A commentary on Bill VanPatten's "Processing instruction: An update". *Language Learning, 52*(4), 805-823.

DeKeyser, R., & Sokalski, K. (1996). The differential role of comprehension and production practice. *Language Learning, 46*(4), 613-642.

DeKeyser, R., & Sokalski, K. (2001). The differential role of comprehension and production practice. In R. Ellis (Ed.), *Form-focused instruction and second language learning* (pp. 81-112). Malden, MA: Blackwell Publishers.

DfE. (2011). *Review of the national curriculum in England: Summary report of the call for evidence*. London: Department for Education.

DfE. (2013a). *National curriculum in England: English programmes of study (Key stages 1 and 2)*. London: Department for Education.

DfE. (2013b). *The National Curriculum in England: Framework document for teaching 1 September 2014 to 31 August 2015*. London: Department for Education.
DfE. (2013c). *The National Curriculum in England: Key stages 1 and 2 framework document*. London: Department for Education.
DfE. (2013d). *National curriculum in England: Languages programmes of study (Key stage 2)* London: Department for Education.
DfEE, & QCA. (1999). *The national curriculum: Handbook for primary teachers in England (Key Stages 1 and 2)*. London: Department for Education and Employment / Qualifications and Curriculum Authority.
DfES. (2005). *Key stage 2 framework for languages*. London: Department for Education and Skills.
Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development, 79*(4), 1152-1167.
Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Oxford: Blackwell Publishing Ltd.
Doughty, C. J. (2003). Instructed SLA: Constraints, compensation, and enhancement. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 256-310). Oxford: Blackwell Publishing Ltd.
Doughty, C. J. (2004). Commentary: When PI is focus on form it is very, very good, but when it is focus on forms... In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 272-270). New Jersey: Lawrence Erlbaum Associates, Inc.
Doughty, C. J., & Williams, J. (1998a). Issues and terminology. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 1-12). Cambridge: Cambridge University Press.
Doughty, C. J., & Williams, J. (1998b). Pedagogical choices in focus on form. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 197-262). Cambridge: Cambridge University Press.
Dreyer, H., & Schmitt, R. (2001). *A practice grammar of German*. Ismaning: Max Hueber Verlsg.
Driscoll, P., Jones, J., & Macrory, G. (2004). *The provision of foreign language learning for pupils at key stage 2 (DfES Research Report 572)*. London: Department for Education and Skills.
Ellis, N. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*(2), 143-188.
Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*(3), 305-352.
Ellis, N. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164-194.
Ellis, N. (2010). Optimizing the input: Frequency and sampling in usage-based and form-focused learning. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 139-158). Chichester: Wiley-Blackwell.
Ellis, N., & Sagarra, N. (2011). Learned attention in adult language acquisition: A replication and generalization study. *Studies in Second Language Acquisition, 33*(4), 589-624.
Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.
Ellis, R. (1999). Input-based approaches to teaching grammar: A review of classroom-oriented research. *Annual Review of Applied Linguistics, 12*(1), 64-90.
Ellis, R. (2001). Investigating form-focused instruction. In R. Ellis (Ed.), *Form-focused instruction and second language learning* (pp. 1-46). Malden, MA: Blackwell Publishers.
Ellis, R. (2002). Does form-focussed instruction affect the acquisition of implicit knowledge? *Studies in Second Language Acquisition, 24*(2), 223-236.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54*(2), 227-275.
Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. *Studies in Second Language Acquisition, 27*(2), 141-172.
Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly, 40*(1), 83-107.
Ellis, R. (2009a). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-26). Bristol: Multilingual Matters.
Ellis, R. (2009b). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31-64). Bristol: Multilingual Matters.
Ellis, R. (2009c). Retrospect and prospect. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing, and teaching* (pp. 335-353). Bristol: Multilingual Metters.
Ellis, R., & Shintani, N. (2013). *Exploring language pedagogy through second language acquisition research*. London: Routledge.
Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*(3), 464-491.
Erlam, R. (2009). The elicited oral imitation test as a measure of implicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 65-93). Bristol: Multilingual Matters.
Farley, A. P. (2004a). Processing instruction and the Spanish subjunctive: Is explicit information needed? In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 227-239). New Jersey: Lawrence Erlbaum Associates, Inc.
Farley, A. P. (2004b). The relative effects of processing instruction and meaning-based output instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 143-168). New Jersey: Lawrence Erlbaum Associates, Inc.
Felix, S. W. (1985). More evidence on competing cognitive systems. *Second Language Research, 1*(1), 47-72.
Fernandez, C. (2008). Re-exmaining the role of explicit information in processing instruction. *Studies in Second Language Acquisition, 30*(3), 277-305.
Field, A. (2009). *Discovering statistics using SPSS* (Third ed.). London: Sage.
Gass, S., & Mackey, A. (2007). *Data elicitation for second and foreign language research*. New Jersey: Lawrence Erlbaum Associates.
Gass, S., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed.). New Jersey: Lawrence Erlbaum Associates.
Goldschneider, J. M., & DeKeyser, R. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning, 51*(1), 1-50.
Graham, S., Courtney, L., Marinis, T., & Tonkyn, A. (2014). *Primary modern languages: The impact of teaching approaches on attainment and preparedness for secondary school language learning (Final report)*. London: Nuffield Foundation.
Green, P. S., & Hecht, K. (1992). Implicit and explicit grammar: An empirical study. *Applied Linguistics, 13*(2), 168-184.
Grey, S., Rebuschat, P., & Williams, J. (2014). Incidental exposure and L3 learning of morphosyntax. *Studies in Second Language Acquisition, 36*(4), 611-645.
Hama, M., & Leow, R. P. (2010). Learning without awareness revistied: Extending Williams 2005. *Studies in Second Language Acquisition, 32*(3), 465-491.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research, 2*(1), 1-23.

Hanan, R. E. (2011). *Explicit knowledge about language in primary school children: Exploring the potential effectiveness of grammar instruction in primary school level foreign language learning*. Unpublished MA dissertation. Department of Education. University of York. York.

Harley, B. (1998). The role of focus-on-form tasks in promoting child L2 acquisition. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition.* (pp. 156-174). Cambridge: Cambridge University Press.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19*(3), 379-400.

Harley, B., & Hart, D. (2002). Age, aptitude and second language learning on a bilingual exchange. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 301-330). Amsterdam: John Benjamins Publishing Company.

Hawkins, E. (1999). Foreign language study and language awareness. *Language Awareness, 8*(3), 124-142.

Hawkins, E. (2005). Out of this nettle, drop out, we pluck this flower, opportunity: Re-thinking the school foreign language apprenticeship. *Language Learning Journal, 32*(1), 4-17.

Henry, N., Culman, H., & VanPatten, B. (2009). More on the effects of explicit information in instructed SLA: A partial replication and a response to Fernández (2008). *Studies in Second Language Acquisition, 31*(4), 559-575.

Howell, D. C. (2008). *Fundamental statistics for the behavioral sciences* (6th ed.). Belmont, CA: Thomson Wadsworth.

Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). California: Wadsworth.

Hulstijn, J. H. (1997). Second language acquisition research in the laboratory: Possibilities and limitations. *Studies in Second Language Acquisition, 19*(1), 131-143.

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing, and acquisition of second language knowledge. *Second Language Research, 18*(3), 193-223.

Hulstijn, J. H., & de Graaf, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11*(1), 97-112.

Hunt, M., Barnes, A., Powell, B., Lindsay, G., & Muijs, D. (2005). Primary modern foreign languages: An overview of recent reserach, key issues and challenges for educational policy and practice. *Research Papers in Education, 20*(4), 371-390.

Hunt, M., Barnes, A., Powell, B., Lindsay, G., & Mujis, D. (2005). Primary modern foreign languages: An overview of recent research, key issues, and challenges for educational policy and practice. *Research Papers in Education, 20*(4), 371-390.

Hurry, J., Nunes, T., Bryant, P., Pretzlik, U., Parker, M., Curno, T., & Midgely, L. (2005). Transforming research on morphology into teacher practice. *Research Papers in Education, 20*(2), 187-206.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539-587). Oxford: Blackwell Publishing Ltd.

Isemonger, I. M. (2007). Operational definitions of explicit and implicit knowledge: Response to R. Ellis (2005) and some recommendations for future research in this area. *Studies in Second Language Acquisition, 29*(1), 101-118.

Issidorides, D. C., & Hulstijn, J. H. (1992). Comprehension of grammatically modified and nonmodified sentences by second language learners. *Applied Psycholinguistics, 13*(2), 147-171.

Jackson, C. (2007). The use and non-use of semantic information, word order and case-markings during comprehension by L2 learners of German. *The Modern Language Journal, 91*(3), 418-432.

Johnson, J. S. (1992). Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence. *Language Learning, 42*(2), 217-248.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.

Jourdenais, R., Ota, M., Stauffer, S., Boyson, B., & Doughty, C. J. (1995). Does textual enhancment promote noticing? A think-aloud protocol analysis. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 183-216). Hawaii: Hawaii University Press.

Kellerman, E. (1985). If at first you do succeed... In S. M. Gass & C. Madden (Eds.), *Input and second language acquisition* (pp. 345-353). Rowley, MA: Newbury House.

Kempe, V., & MacWhinney, B. (1998). The acquisition of case-marking by adult learners of Russian and German. *Studies in Second Language Acquisition, 20*(4), 543-587.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon Press Ltd.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. New York: Prentice Hall.

Krashen, S. D., Long, M. H., & Scarcella, R. C. (1979). Age, rate, and eventual attainment in second language acquisition. *TESOL Quarterly, 13*(4), 573-582.

Lanzer, H., & Wardle, M. (2009). *Edexcel GCSE German*. Essex: Pearson Education Limited.

Larsen-Freeman, D. (1997). Chaos / complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141-165.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619.

Larsen-Freeman, D. (2014). Complexity Theory: Renewing Our Understanding of Language, Learning, and Teaching. Keynote Speech at TESOL 2014: International conference and English language expo. Portland, Oregan. Retrieved January 20, 2015, from http://www.tesol.org/attend-and-learn/international-convention/convention2014/featured-speakers/diane-larsen-freeman-keynote-video

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Laval, C. (2013). The age factor on the primary, secondary and cumulative transfer-of-training effects of processing instruction on the acquisition of French as a second language. In J. F. Lee & A. Benati (Eds.), *Individual differences and processing instruction* (pp. 105-130). Sheffield: Equinox Publishing.

Lee, J. F. (1987). Morphological factors influencing pronominal reference assignment by learners of Spanish. In T. A. Morgan, J. F. Lee & B. VanPatten (Eds.), *Language and language use: Studies in Spanish* (pp. 221-232). Lanham, MD: University Press of America.

Lee, J. F., & Benati, A. (2007). *Second language processing: An analysis of theory, problems, and possible solutions*. London: Continuum.

Lee, J. F., & Malovrh, P. A. (2009). Linguistic and non-linguistic factors affecting OVS processing of accusative and dative case pronouns by advanced L2 learners of Spanish. In J. Collentine (Ed.), *Selected proceedings of the 11th Hispanic Linguistics Symposium* (pp. 105-116). Somerville, MA: Cascadilla Proceedings Project.

Lee, S.-K., & Huang, H.-T. (2008). Visual input enhancement and grammar learning. *Studies in Second Language Acquisition, 30*(3), 307-331.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: John Wiley and Sons.

Leow, R. P. (2001). Attention, awareness, and foreign language behavior. In R. Ellis (Ed.), *Form-focused instruction and second language learning* (pp. 113-155). Oxford: Blackwell Publishers.

Leow, R. P. (2015). The role of attention and (un)awareness in SLA: Conceptual replication of N. C. Ellis and Sagarra (2010a) Leung and Williams (2012). *Language Teaching, 48*(1), 117-129.

Leung, J. H. C., & Williams, J. (2011). The implicit learning of mappings between forms and contextually derived meanings. *Studies in Second Language Acquisition, 33*(1), 33-55.

Leung, J. H. C., & Williams, J. (2012). Constraints on implicit learning of grammatical form-meaning connections. *Language Learning, 62*(2), 634-662.

Lightbown, P. M. (2000). Classroom SLA research and second language teaching. *Applied Linguistics, 21*(4), 431-462.

LoCoco, V. (1987). Learning comprehension of oral and written sentences in German and Spanish: The importance of word order. In B. VanPatten, T. Dvorak & J. F. Lee (Eds.), *Foreign language learning: A research perspective* (pp. 119-129). Rowley, MA: Newbury House.

Loewen, S., Erlam, R., & Ellis, R. (2009). The incidental acquisition of third person -s as implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 262-280). Bristol: Multilingual Matters.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition, 12*(3), 251.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg & C. Kramsch (Eds.), *Foreign language research in cross-cultural persepctive* (pp. 39-52). Amsterdam: John Benjamins.

Long, M. H. (2010). Methodological principles for language teaching. In M. H. Long & C. J. Doughty (Eds.), *The handbook of language teaching* (pp. 373-394). Chichester: Wiley-Blackwell.

Long, M. H., & Robinson, P. (1998). Focus on form: Theory, research, and practice. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 15-41). Cambridge: Cambridge University Press.

Loschky, L., & Bley-Vroman, R. (1993). Grammar and task-based methodology. In G. Crookes & S. Gass (Eds.), *Tasks and language learning* (pp. 123-167). London: Multilingual Matters.

Mackey, A., & Gass, S. (2005). *Second language research: Methodology and design*. New Jersey: Lawrence Erlbaum Associates, Inc.

MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249-308). New Jersey: Lawrence Erlbaum.

MacWhinney, B. (2001). The competition model: The input, the context, and the brain. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 69-90). New York: Cambridge University Press.

MacWhinney, B. (2005). A unified model of language acquisition. In J. Kroll & A. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49-67). Oxford: Oxfrod University Press.

MacWhinney, B., & Bates, E. (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior, 23*(2), 127-150.

Marchman, V., & Thal, D. (2005). Words and grammar. In M. Tomasello & D. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates*. New Jesery: Lawrence Erlbaum Associates, Inc.

Marsden, E. (2006). Exploring input processing in the classroom: An experimental comparison of processing instruction and enriched input. *Language Learning, 56*(3), 507-566.

Marsden, E. (2007). Can educational experiments both test a theory and inform practice? *British Educational Research Journal, 33*(4), 565-588.

Marsden, E., Altmann, G., & St Claire, M. (2013). Priming of verb inflections in L1 and L2 French: A comparison of 'redundant' and 'non-redundant' training conditions. *International Review of Applied Linguistics, 51*(3), 271-298.

Marsden, E., & Chen, H.-Y. (2011). The roles of structured input activities in processing instruction and the kinds of knowledge they promote. *Language Learning, 61*(4), 1058-1098.

Marsden, E., & Torgerson, C. (2012). Single group, pre- and post-test research designs: Some methodological concerns. *Oxford Review of Education, 5*, 583-616.

Martin, C. (2000). Modern foreign languages at primary school: A three-pronged approach? *Language Learning Journal, 22*(1), 5-10.

Mavrantoni, M., & Benati, A. (2013). The effects of processing instruction and traditional instruction on two different school-age learners: The case of English present simple tense third person singular. In J. F. Lee & A. Benati (Eds.), *Individual differences and processing instruction* (pp. 185-209). Sheffield: Equinox Publishing.

Milton, J., & Alexiou, T. (2006). Language aptitude development in young learners. In C. Abello-Contesse, R. Chacón-Beltrán, M. D. López-Chiménez & M. M. Torreblanca-López (Eds.), *Age in L2 acquisition and teaching* (pp. 177-192). Oxford: Peter Lang.

Minette, E., Crellin, K., & Holden, C. (2009). French schemes of work. Retrieved February 11, 2015, from https://www.westsussex.gov.uk/learning/west_sussex_grid_for_learning/curriculum/modern_foreign_languages_mfl/key_stage_2/french_schemes_of_work.aspx

Mitchell, R. (2011). Still gardening in a gale: Policy, research and practice in foreign language education in England. *Fremdsprachen Lehren und Lernen, 40*(1), 49-67.

Mitchell, R., Myles, F., & Marsden, E. (2012). *Second language learning theories* (Third ed.). Oxon: Routledge.

Morgan-Short, K., & Bowden, H. W. (2006). Processing instruction and meaningful output-based instruction: Effects on second language development. *Studies in Second Language Acquisition, 28*(1), 31-65.

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning, 60*(1), 154-193.

Muijs, D., Barnes, A., Hunt, M., Powell, B., Arweck, E., Lindsay, G., & Martin, C. (2005). *Evaluation of the key stage 2 language learning pathfinders (DfES Research Report 692)*. London: Department for Education and Skills.

Muñoz, C. (2006). The effects of age on foreign language learning: The BAF project. In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 1-40). Clevedon: Multilingual Matters.

Muñoz, C. (2007). Age-related differences and second language learning practice. In R. Dekeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 229-255). Cambridge: Cambridge University Press.

Muñoz, C. (2008a). Age-related differences in foreign language learning: Revisiting the empirical evidence. *International Review of Applied Linguistics, 46*(3), 197-220.

Muñoz, C. (2008b). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics, 29*(4), 578-596.

Muñoz, C. (2014). Exploring young learners' foreign language learning awareness. *Language Awareness, 23*(1), 24-40.

Muñoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching, 44*(1), 1-35.

Myles, F., & Mitchell, R. (2012). *Learning French from ages 5, 7 and 11: An investigation into starting ages, rates and routes of learning amongst early foreign language learners (End of award report, RES-062-23-1545)*. Swindon: ESRC.

Neubauer, K., Pearson, N., & Whittle, E. (2015). *Deutsch mit Felix und Franzi: New lesson plans for German at key stage 2 (volume 2)*. London: Goethe-Institut.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*(1), 11-28.

Nicholas, H., & Lightbown, P. M. (2008). Defining child second language acquisition, defining roles for L2 instruction. In J. Philp, R. Oliver & A. Mackey (Eds.), *Second language acquisition and the younger learner: Child's play?* (pp. 27-52). Amsterdam: John Benjamins Publishing Company.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417-528.

Nunan, D. (1991). Methods in second language classroom-oriented research. *Studies in Second Language Acquisition, 13*(2), 249-274.

Nunes, T., Bryant, P., & Olsson, J. (2003). Learning morphological and phonological spelling rules: An intervention study. *Scientific Studies of Reading, 7*(3), 289-307.

OFSTED. (2005). *Implementing languages entitlement in primary schools*. London: OFTSED.

Paradis, J. (2007). Second language acquisition in childhood. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development*. Malden MA: Blackwell.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393-419). San Diego: Academic Press.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. New York: Atheneum.

Philp, J., Mackey, A., & Oliver, R. (2008). Child's play? Second language acquisition and the younger learner in context. In J. Philp, R. Oliver & A. Mackey (Eds.), *Second language acquisition and the younger learner: Child's play?* (pp. 3-26). Amsterdam: John Benjamins Publishing Company.

Philp, J., & Tognini, R. (2009). Language acquisition in foreign language contexts and the differential benefits of interaction. *International Review of Applied Linguistics, 47*(2), 245-266.

Pienemann, M. (1989). Is language teachable? Psycholinguistc experiments and hypotheses. *Applied Linguistics, 10*(1), 52-79.

Pimsleur, P. (1966). *The Pimsleur language aptitude battery*. New York: Harcourt, Brace, Jovanovic.

QCA. (2007a). Key stage 2 scheme of work for languages: French. Retrieved February 9, 2015, from http://webarchive.nationalarchives.gov.uk/20090318150633/qca.org.uk/qca_11752.aspx#French

QCA. (2007b). Key Stage 2 scheme of work for languages: German. Retrieved February 9, 2015, from http://webarchive.nationalarchives.gov.uk/20090318150633/qca.org.uk/qca_11752.aspx#German

QCA. (2007c). Key stage 2 scheme of work for languages: Spanish. Retrieved February 9, 2015, from http://webarchive.nationalarchives.gov.uk/20090318150633/qca.org.uk/qca_11752.aspx#Spanish

Ranta, L. (2002). The role of learners' language analytic ability in the communicative classroom. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 159-180). Amsterdam: John Benjamins Publishing Company.

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modelling and Analytics, 2*(1), 21-33.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning, 63*(3), 595-626.

Rebuschat, P., & Williams, J. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33*(4), 829-856.

Reinders, H., & Ellis, R. (2009). The effects of two types of input on intake and the acquisition of implicit and explicit knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 281-302). Bristol: Multilingual Matters.

Robinson, P. (1995). Aptitude, awareness and the fundamental similarity of implicit and explicit second language learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 303-358). Honolulu: University of Hawai'i at Manoa.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*(1), 45-99.

Robinson, P. (2002). Learning conditions, aptitude complexes and SLA: A framework for research and pedagogy. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 113-133). Amsterdam: John Benjamins Publishing Company.

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 631-678). Oxford: Blackwell Publishing Ltd.

Robson, C. (2011). *Real world research.* Chichester: Wiley.

Roehr, K. (2006). Metalinguistic knowledge and language-analytic ability in university-level L1 learners. *Essex Research Reports in Linguistics, 51*(1), 41-71.

Roehr, K. (2008). Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics, 29*(2), 173-199.

Roehr, K. (2012). The Springboard to Languages evaluation project: A summary report. In A. Tellier (Ed.), *Esperanto as a starter language for child second-language learners in the primary school* (pp. 23-34). Stoke-on-Trent: Esperanto UK.

Rosa, E. M., & Leow, R. P. (2004a). Awareness, different learning conditions, and second language development. *Applied Psycholinguistics, 25*(2), 269-292.

Rosa, E. M., & Leow, R. P. (2004b). Computerized task-based exposure, explicitness, type of feedback, and Spanish L2 development. *The Modern Language Journal, 88*(2), 192-216.

Rose, J. (2009). *Independent review of the primary curriculum: Final report*. London: Crown Copyright.

Russell, V. (2012). Learning complex grammar in the virtual classroom: A comparison of processing instruction, structured input, computerized visual enhancement, and traditional instruction. *Foreign Language Annals, 45*(1), 42-71.

Sanz, C. (2004). Computer delivered implicit versus explicit feedback in processing instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 241-256). New Jersey: Lawrence Erlbaum Associates, Inc.

Sanz, C., & Morgan-Short, K. (2004). Positive evidence versus explicit rule presentation and explicit negative feedback: A computer-assisted study. *Language Learning, 54*(1), 35-78.

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 319-353). Cambridge: Cambridge University Press.

Schicker, C., Waltl, M., & Malz, C. (2011). *Zoom Deutsch 1*. Oxford: Oxford University Press.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158.

Schmidt, R. (1994). Implicit learning and the cognitive unconscious: Of artificial grammars and SLA. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 165-209). London: Academic Press.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1-63). Hawaii: Hawaii University Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmidt, R., & Frota, S. (1986). Developing basic conversational ability in a second language: A case study of an adult learner of Portuguese. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 237-326). Rowley, MA: Newbury House.

Seccombe, C. (2014a). *French minibooks for primary learners.* Ideas Education Ltd.

Seccombe, C. (2014b). Primary French resources. Retrieved February 11, 2015, from http://www.lightbulblanguages.co.uk/resources-pr-fr-resources.htm#gen

Seccombe, C. (2014c). *Spanish minibooks for primary learners*. Ideas Education Ltd.

Shak, J., & Gardner, S. (2008). Young learner perspecitves on four focus-on-form tasks. *Language Teaching Research, 12*(3), 387-408.

Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research, 7*(2), 118-132.

Sharwood Smith, M. (1993). Input enhancement and instructed second language acquisition. *Studies in Second Language Acquisition, 15*(1), 65-85.

Simard, D. F. (2009). Differential effects of textual enhancement formats on intake. *System, 37*(1), 124-135.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxofrd: Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-94). Amsterdam: John Benjamins Publishing Company.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development, 49*(4), 1114-1128.

Spada, N. (1997). Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching, 30*(2), 73-87.

Spada, N. (2005). Conditions and challenges in developing school-based SLA research programs. *The Modern Language Journal, 89*(3), 328-338.

Spada, N., & Lightbown, P. M. (1993). Instruction and the development of questions in L2 classrooms. *Studies in Second Language Acquisition, 15*(2), 205-224.

Spada, N., Lightbown, P. M., & White, J. (2005). The importance of form/meaning mappings in explicit form-focused instruction. In A. Housen & M. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 199-234). Berlin: Mouton de Gruyer.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*(2), 263-308.

Stafford, C. A., Bowden, H. W., & Sanz, C. (2012). Optimizing language instruction: Matters of explicitness, practice and cue learning. *Language Learning, 62*(3), 741-768.

Stevens, J. P. (2002). Applied multivariate statistics for the social sciences. New Jersey: Erlbaum.

Sukhnandan, L., & Lee, B. (1998). *Streaming, setting and grouping by ability: A review of the literature*. National Foundation for Educational Research.

Svalberg, A. M.-L. (2012). Language awareness in language learning and teaching: A research agenda. *Language Teaching, 45*(3), 376-388.

Swain, M. (1998). Focus on form through conscious reflection. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64-82). Cambridge: Cambridge University Press.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon.

Tellier, A. (2012a). Esperanto as a potential aid to language learning in primary schools. In A. Tellier (Ed.), *Esperanto as a starter language for child second-langauge learners in the primary schools* (pp. 10-18). Stoke-on-Trent: Esperanto UK.

Tellier, A. (2012b). *Esperanto as a starter language for child second-language learners in the primary school*. Stoke-on-Trent: Esperanto UK.

Tellier, A., & Roehr-Brackin, K. (2013a). The development of language learning aptitude and metalinguistic awareness in primary-school children: A classroom study. *Essex Research Reports in Linguistics, 62*(1), 1-28.

Tellier, A., & Roehr-Brackin, K. (2013b). Metalinguistic awareness in children with differing language learning experience. *EUROSLA Yearbook, 13*, 81-108.

The Nuffield Languages Inquiry. (2000). *Languages: The next generation*. London: The Nuffield Foundation.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*(3), 209-253.

Tomita, Y., Suzuki, W., & Jessop, L. (2009). Elicited imitation: Toward valid procedures to measure implicit second language grammatical knowledge. *TESOL Quarterly, 43*(2), 345-350.

Tomlin, R. S., & Villa, V. (1994). Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition, 16*, 185-204.

Torgerson, C. J., & Torgerson, D. J. (2001). The need for randomised controlled trials in educational research. *British Journal of Education Studies, 49*(3), 316-328.

Trahey, M., & White, L. (1993). Positive evidence and preemption in the second language classroom. *Studies in Second Language Acquisition, 15*(2), 181-204.

Truscott, J. (1998). Noticing in second language acquisition: A critical review. *Second Language Research, 14*(2), 103-135.

Truscott, J. (1999). What's wrong with oral grammar correction. *Canadian Modern Language Review, 55*(4), 437-456.

Truscott, J. (2004). The effectiveness of grammar instruction: Analysis of a meta-analysis. *English Teaching and Learning, 28*(3), 17-29.

Truscott, J. (2005). The continuing problems of oral grammar correction. *The International Journal of Foreign Language Teaching, 1*(2), 17-25.

VanPatten, B. (1984). Learners' comprehension of clitic pronouns: More evidence for a word order strategy. *Hispanic Linguistics, 1*, 57-67.

VanPatten, B. (1996). *Input processing grammar instruction: Theory and research*. New Jersey: Ablex Publishing Corporation.

VanPatten, B. (2002). Processing instruction: An update. *Language Learning, 52*(4), 755-803.

VanPatten, B. (2004a). Input processing in SLA. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5-32). New Jersey: Lawrence Erlbaum Associates, Inc.

VanPatten, B. (2004b). Processing instruction versus other types of instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 93-96). New Jersey: Lawrence Erlbaum Associates, Inc.

VanPatten, B. (2004c). *Processing instruction: Theory, research, and commentary*. New Jersey: Lawrence Erlbaum Associates, Inc.

VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 115-136). New Jersey: Lawrence Erlbaum Associates, Inc.

VanPatten, B. (2014). The psycholinguistic of SLA. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 1-19). New York: Routledge.

VanPatten, B., & Borst, S. (2012). The role of explicit information and grammatical sensitivity in processing instruction: Nominative-accusative case marking and word order in German L2. *Foreign Language Annals, 45*(1), 92-109.

VanPatten, B., Borst, S., Collopy, E., & Qualin, A. (2013). Explicit information, grammatical sensitivity, and the first noun principle: A cross-linguistic study in processing instruction. *The Modern Language Journal, 97*(2), 506-527.

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition, 15*(2), 225-243.

VanPatten, B., & Fernández, C. (2004). The long-term effects of processing instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 273-290). New Jersey: Lawrence Erlbaum Associates, Inc.

VanPatten, B., & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition, 18*(4), 495-510.

VanPatten, B., & Sanz, C. (1995). From input to output: Processing instruction and communicative tasks. In F. Eckman, D. Highland, P. Lee, J. Mileham & R. Rutowski Weber (Eds.), *Second language acquisition theory and pedagogy*. New Jersey: Erlbaum.

VanPatten, B., & Wong, W. (2004). Processing instruction and the French causative: Another replication. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 97-118). New Jersey: Lawrence Erlbaum Associates, Inc.

Wade, P., Marshall, H., & O'Donnell, S. (2009). *Primary modern foreign languages: Longitudinal survey of implementation of national entitlement to language learning at Key Stage 2 (Final report)*. London: National Foundation for Education Research.

Weinberg, S. L., & Abramowitz, S. K. (2002). *Data analysis for the behavioural sciences using SPSS*. New York: Cambridge University Press.

White, J. (1998). Getting the learners' attention: A typographical input enhancement study. In C. J. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 53-113). Cambridge: Cambridge University Press.

White, J. (2008). Speeding up acquisition of *his* and *her*: Explicit L1/L2 contrasts help. In J. Philp, R. Oliver & A. Mackey (Eds.), *Second language acquisition and the younger learner: Child's play?* (pp. 193-230). Amsterdam: John Benjamins Publishing Company.

White, J., & Ranta, L. (2002). Examining the interface between metalinguistic task performance and oral production in a second language. *Language Awareness, 11*(4), 259-290.

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research, 12*(3), 233-265.

White, L., Spada, N., Lightbown, P. M., & Ranta, L. (1991). Input enhancement and question formation. *Applied Linguistics, 12*(4), 416-432.

Wilcox, R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist, 53*(3), 300-314.

Williams, J. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*(2), 269-304.

Wong, W. (2004a). The nature of processing instruction. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 33-64). New Jersey: Lawrence Erlbaum Associates, Inc.

Wong, W. (2004b). Processing instruction in French: The roles of explicit information and structured input. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 187-206). New Jersey: Lawrence Erlbaum Associates, Inc.