# The effects of facilitated feedback on the second-language English writing of Korean university students

Jill Boggs

Linacre College

Thesis submitted to the University of Oxford for the degree of Doctor of Philosophy

Hilary Term 2018

Words: 96,946

# Acknowledgements

There are more people to thank than there is space to thank them, but a few do deserve special mention. I thank my supervisors, Catherine Walter and Robert Woore, for the time you spent reading and commenting on my many drafts, for your ideas and direction, and for insisting that I give my best. I thank Catherine for getting me started on this journey with the MSc course!

To the two very dear friends and colleagues who opened their classrooms to me: this project would never have happened without you. Thank you for your generosity with your time, your syllabus, and your classrooms. Thank you for making time to see me when I return to Korea for visits. Thank you to other colleagues as well – for piloting my instruments, for sharing your office space while I collected data, and for your friendship. To the students who graciously agreed to participate in this project and share your language learning journey with me: thank you.

To the librarians of the Education Library: Thank you for your friendly smiles, and willingness to help with pretty much anything, from the mundane (gaiters!) to doggedly tracking down sources I needed. Thank you for letting me sit in reception and chat with you about nothing at all. The emotional break away from my work kept me sane. Thank you for having vegan chocolate at the desk at Christmas one year – made my day!

Thank you to Ross and Howard in Linacre IT for being amazing and for maintaining library computers. Thank you for going above and beyond to offer help solving problems. Thank you to Nick B. and admin staff for being friendly and helpful; to Michael and the Dining hall staff for sustenance and smiles – a special thanks to the staff for the hugs! To Aiga and Housekeeping for smiles, hugs, and a pleasant working environment. Thank you to Nick D. and the maintenance team, who ensured our safety and comfort and went above and beyond the call of duty – thank you. The staff at Linacre have given up evenings, weekends, and holidays to make things work here. Thank you for this and for helping me in a hundred other ways that made a big difference. Your smiles and friendly greetings were sometimes the only bright spots in the day, so thank you for doling them out so generously! Thank you to Jane, who added me to her already-lengthy to-do list when my advisor moved away. Thank you for your time, your amazing care, and our many oat milk lattes.

Thank you to Richard Pring, who lent me the resource that made some things snap into place after months of searching. Thank you for being a reliably cheery entity in what was otherwise sometimes a very dark world. Thank you for caring. Thank you to Jill and John Fresen for your friendship and encouragement. It's good to know there are academics like you in the world; rest in peace, John. Thank you to David Langslow, who generously offered his time to help me with an instrument, and to Laerd Statistics for the time they spent answering my questions.

Thank you to my CoNi gals, who generously offered help even when they had their own work to do, and without whose unfading support, love, humour, and encouragement my soul could not have survived intact. Thank you for offering me your friendship when I was too lost to seek it out. Let us take our message with us as we continue through academia. Thank you to Magda, who made dark, cold, rainy days in Oxford not only bearable, but cozy. Thanks for commiseration and for coffee, and for mulled wine and port (whatever that is!). Thanks to Yasmine, who has been a much better friend than I've deserved - thank you for tea, laughs, love, and encouragement. Thank you to Sarah, who always rearranged her life to schedule time with me when I returned home for a visit – you make me feel loved.

Thank you to my family: to my husband, who took on life as a single guy while I married my thesis, and who welcomed me home with my favourite foods on each visit – 고마워!; to Mom, who raised me with books, letting me borrow as many books from the library as I was able to carry, and who developed my early literacy by reading to me until you were hoarse; thank you for your support and for always being willing to proofread. To my dear Duster, thank you for believing in me, for giggles, for Christmas decorations which are sometimes, but not always, pink, and for more encouragement and kindness than I deserve; to Dad for the kilos and kilos of peanut butter that kept me going; to Aunt Denna, who somehow managed to persuade me on the darkest days that there was light, and – importantly – that one day I would get to see it.

I owe a special measure of gratitude to a dear friend, who stepped in when I was very nearly spent from floundering for far too long and calmed the storm and created a space for me to just *be*. I was too spent to ask for help, too busy floundering to identify what kind of help I might need; you saw what I needed and just stepped in. The space you created allowed me to pause and reflect, and to look around to see where I'd come from and figure out where I needed to go and how I might get there. You restored my faith in my work, encouraging me to see it not as a broken mess, but as a thoughtfully designed creation which had something of value to offer. Your care and kindness rescued my spirit. I can never repay what you have done for me, but I can promise that I will spend the rest of my life paying it forward. Thank you.

# Abstract

Corrective feedback (CF) has been demonstrated in the literature to have a positive effect on language learners' grammatical accuracy in second-language (L2) writing. However, while much work has been done on feedback *type* (e.g. direct and indirect CF), studies continue to produce results which seem to contradict each other, indicating that a different focus might be useful. Moreover, much of the work on CF has been conducted with English majors or with participants studying (or preparing to study) in universities in English-speaking settings, but the generalisability of these studies to settings where English is neither the dominant language nor the learners' principal academic focus is unclear. The current study shifts the focus from type of CF to *ways of facilitating CF use*, exploring the issue with Korean L1 learners of English who are natural science majors at a university in South Korea taking a required English L2 writing course. The current study contributes to the body of knowledge on CF provision by quantitatively and qualitatively investigating how participants' use of feedback in this context is facilitated by their Western teachers, and how the effects of facilitated feedback use compares to effects in a group whose use of feedback is not facilitated.

Working within a cognitive framework of language learning and using a quasi-experimental, mixed-methods, pre-test / post-test / delayed post-test study design, the study explores the differential effects of two ways of facilitating use of CF on the development of grammatical accuracy in written work, which was academic in style. Effects of these two methods of facilitating the use of CF are compared to effects of an unfacilitated condition. The hypothesis that learners would apply explicit knowledge to analyse and address errors in their writing, resulting in increased accuracy in the use of standard forms, is a cognitive one; and the facilitating of the interventions applies concepts from sociocultural theory. Participants in one facilitated condition received CF orally in individual conferences, with the teacher giving contingent assistance and only providing direct CF if needed; participants in the other facilitated condition received direct written CF, and were provided with worksheets designed to structure their independent reflection upon their writing. Finally, the Comparison Group received direct written CF without any facilitation. Measures of the development of accuracy focused on specific grammatical features targeted for instructional and feedback purposes, and assessed improvement in the accuracy in use of those features in new writing. The study applied a process-writing protocol: *Write – Receive CF – Revise – Write a new paragraph*. Effects of self-efficacy and language aptitude on the effectiveness of the treatments are also considered.

Quantitative analyses of the data suggest that providing direct written CF without facilitating its use enabled the development of accuracy equally as well as CF whose use was facilitated, whether by the conferences or the reflective worksheets. Qualitative analyses reveal how teachers facilitated the use of CF in conferences, and how students participated in both forms of facilitation. Findings suggest that previous educational experience may influence learners' ability to effectively use facilitated CF. Finally, interviews with participating teachers and students provide insight into both parties' experiences with these interventions, contributing to knowledge about intervention development and implementation, both in research and in classroom settings.

# Table of Contents

Acknowledgements ..... i
Abstract ..... iii
Table of Contents ..... iv
Tables and figures ..... vii
Abbreviations ..... ix
1 Introduction ..... 1
1.1 Rationale ..... 2
1.2 Background ..... 4
1.2.1 The status of education in Korea ..... 4
1.2.2 The status and teaching of English in Korea ..... 5
1.3 Contribution of the current study ..... 8
1.4 Organisation of thesis ..... 10
2 Literature review ..... 12
2.1 Introduction ..... 12
2.2 Writing and its relevance to L2 learners ..... 12
2.3 Second language development ..... 14
2.3.1 Theoretical discussion of second language development ..... 15
2.3.2 Grammar instruction for language development ..... 26
2.3.3 Section summary ..... 37
2.4 Corrective feedback for L2 writers ..... 38
2.4.1 Operationalising and measuring CF effectiveness ..... 41
2.4.2 Types of written CF ..... 44
2.4.3 Learner-dependent factors influencing CF effectiveness ..... 52
2.4.4 Teacher-controlled factors influencing CF effectiveness ..... 65
2.5 Chapter Summary ..... 78
3 Methodology ..... 80
3.1 Introduction ..... 80
3.2 Aims of research ..... 80
3.3 Methodological orientation ..... 81
3.4 Research design ..... 83
3.4.1 Research questions ..... 85
3.4.2 Operationalising facilitated feedback ..... 86
3.5 Piloting the study ..... 91
3.5.1 Pilot study I (PS-I) ..... 91
3.5.2 Pilot study II (PS-II) ..... 92
3.6 Data collection for the current study ..... 93
3.6.1 Context ..... 93

3.6.2 Creation of intervention groups ..... 97
3.6.3 Instruments ..... 100
3.6.4 Targeted grammar features ..... 119
3.6.5 Procedure ..... 120
3.6.6 The interventions ..... 124
3.7 Data analysis ..... 131
3.7.1 Analyses of data from the intervention ..... 132
3.7.2 Interview transcript analysis ..... 136
3.7.3 Post-hoc conference transcript analysis ..... 139
3.8 Ethical considerations ..... 146
3.8.1 Consent and confidentiality ..... 146
3.8.2 Ensuring no harm to participants ..... 147
4 Results of quantitative analyses ..... 148
4.1 Introduction to quantitative results ..... 148
4.1.1 Research questions to be quantitatively analysed ..... 148
4.2 Quantitative analyses ..... 149
4.2.1 RQ1a Effect of facilitating feedback use on all targeted errors ..... 152
4.2.2 RQ1b Effect of facilitating feedback use on correct article use ..... 154
4.2.3 RQ2 Effect of facilitating feedback use on self-efficacy development ..... 156
4.2.4 RQ3a: Predictive value of LAA and SE for all targeted errors ..... 157
4.2.5 RQ3b: Predictive value of LAA and SE on correct article use ..... 158
4.2.6 RQ4a: Effect of worksheet quality on all targeted errors ..... 159
4.2.7 RQ4b: Effect of worksheet quality on correct article use ..... 160
4.3 Chapter summary ..... 160
5 Results of post-hoc analyses ..... 162
5.1 Introduction ..... 162
5.2 Conferences ..... 163
5.2.1 What was the nature of the teachers' contributions to the conferences? ..... 163
5.2.2 What is the nature of students' participation in the conferences? ..... 177
5.3 Reflective worksheets ..... 184
5.3.1 What was the nature of student participation with the worksheets? ..... 185
5.4 Comparison of student participation in the two interventions ..... 190
5.5 Chapter summary ..... 192
6 Results of interview analysis ..... 193
6.1 Introduction ..... 193
6.2 Student participant interviews ..... 193
6.2.1 Student participant perceptions of the interventions ..... 196
6.2.2 Student participant perceptions of the class ..... 200

6.2.3 Student participant perceptions of prior learning experiences ..... 208
6.3 Teacher participant interviews ..... 211
6.3.1 Experience with interventions in the current study ..... 212
6.3.2 Past feedback methods ..... 215
6.4 Chapter summary ..... 217
7 Discussion and Conclusions ..... 218
7.1 Introduction ..... 218
7.2 Discussion of research questions ..... 218
7.2.1 Differential effects of facilitating feedback use on outcomes (RQs 1-2) ..... 219
7.2.2 Mediating variables which might influence outcomes (RQs 3-4) ..... 224
7.2.3 RQs 6-9 Post-hoc research questions ..... 227
7.3 Implications for research and pedagogy ..... 235
7.4 Strengths and limitations of the study ..... 236
7.5 Future research ..... 237
7.6 Conclusion ..... 238
References ..... 241
Appendices ..... 271

# Tables and figures


TABLE 2-1 CHARACTERISTICS OF SUZUKI AND DEKEYSER'S (2017) TESTS (REPRODUCED FROM SUZUKI AND DEKEYSER, 2017:761)31
TABLE 2-2 TERMINOLOGY RELATED TO *EXPLICIT KNOWLEDGE* OF LANGUAGE ..... 33
TABLE 2-3 EXPLANATION OF TERMS USED WHEN DISCUSSING ASPECTS OF EXPLICIT KNOWLEDGE ..... 34
TABLE 2-4 WRITTEN AND ORAL CF STRATEGY TAXONOMY (REPRODUCED FROM ELLIS, 2009:8) ..... 41
TABLE 2-5 SUMMARY OF TERMS AND EXPLANATION OF USAGE ..... 68
TABLE 3-1 INTERVENTION TIMETABLE ..... 85
TABLE 3-2 SUMMARY OF INTERVENTION GROUPS AND CORRESPONDING THEORETICAL FRAMEWORKS ..... 91
TABLE 3-3 A COMPARISON OF CREDITS REQUIRED OF A MAINSTREAM HIGH SCHOOL AND A SCIENCE HIGH SCHOOL IN SEOUL ..... 97
TABLE 3-4 DISTRIBUTION OF PARTICIPANTS (N=122) IN INTERVENTION GROUPS BY LEARNING COMMUNITY (LC) ..... 99
TABLE 3-5 GENDER BY INTERVENTION GROUP ..... 99
TABLE 3-6 AGE OF COMMENCING ENGLISH STUDY BY INTERVENTION GROUP ..... 99
TABLE 3-7 KNOWLEDGE OF A FOREIGN LANGUAGE OTHER THAN ENGLISH BY INTERVENTION GROUP ..... 99
TABLE 3-8 PARTICIPANTS ATTENDING SCIENCE HIGH SCHOOLS (OR BEING ON A SCIENCE TRACK IN STANDARD HIGH SCHOOLS) AND THOSE ATTENDING A STANDARD HIGH SCHOOL BY INTERVENTION GROUP ..... 99
TABLE 3-9 NUMBER OF EXTENDED-LENGTH[B] KOREAN L1 ESSAYS STUDENTS REPORT HAVING WRITTEN DURING HIGH SCHOOL CAREER BY INTERVENTION GROUP ..... 100
TABLE 3-10 MEAN L2 PROFICIENCY SCORES (OXFORD ONLINE PLACEMENT TEST) BY INTERVENTION GROUP ..... 100
TABLE 3-11 INSTRUMENTS USED IN THE STUDY WITH AN EXPLANATION OF PURPOSE AND, WHERE RELEVANT, CORRESPONDING RESEARCH QUESTION (RQ) ..... 100
TABLE 3-12 GRAMMAR PRE-TEST SCORES BY GROUP (OUT OF 13) ..... 118
TABLE 3-13 SUMMARY OF THE STUDY ..... 123
TABLE 3-14 SUMMARY OF THE THREE INTERVENTIONS WITH EXPLANATIONS ..... 125
TABLE 3-15 TABLE RELATING RESEARCH QUESTIONS TO INSTRUMENTS AND PLAN OF ANALYSIS ..... 131
TABLE 3-16 CHOICES RESEARCHERS MAKE IN THEMATIC ANALYSIS (BRAUN & CLARKE 2006) ..... 137
TABLE 3-17 MODIFIED CODING SCHEME FOR THE CURRENT STUDY ..... 144
TABLE 4-1 MEAN NUMBERS OF WORDS WRITTEN AND ERRORS (TOTAL AND ARTICLE-ONLY) MADE FOR PRE-, POST-, AND DELAYED POST-TESTS FOR THE THREE GROUPS ..... 150
TABLE 4-2 SHAPIRO-WILK TEST OF NORMALITY FOR RQ1A ..... 152
TABLE 4-3 T-TEST RESULTS FOR PRE-TEST TO POST-TEST AND PRE-TEST TO DELAYED POST-TEST FOR THE THREE GROUPS ..... 154
TABLE 4-4 MEAN PERCENTAGE AND SD FOR CORRECT ARTICLE USES WITH REGARD TO MAXIMUM POSSIBLE CORRECT USES, BY GROUP AND TIME (FIGURES PROVIDED FOR NON-TRANSFORMED DATA) ..... 155
TABLE 4-5 SHAPIRO-WILK TEST OF NORMALITY FOR RQ1B ..... 155
TABLE 4-6 MEAN SELF-EFFICACY SCORES FOR THE INITIAL AND FINAL ADMINISTRATIONS AND THE GAIN SCORES FOR EACH GROUP . 156
TABLE 4-7 LANGUAGE ANALYTIC ABILITY AND SELF-EFFICACY SCORES (MEAN, STANDARD DEVIATION, AND RANGE) ..... 157
TABLE 4-8 REGRESSION COEFFICIENTS AND STANDARD ERRORS FOR RQ3A ..... 158
TABLE 4-9 REGRESSION COEFFICIENTS AND STANDARD ERRORS FOR RQ3B ..... 159
TABLE 4-10 MEAN WORKSHEET QUALITY PERCENTAGE SCORES BY ITERATION AND BY TEACHER (WITH STANDARD DEVIATION AND RANGE) ..... 159
TABLE 5-1 MEAN LENGTH IN MINUTES OF CONFERENCES (WITH STANDARD DEVIATION AND RANGE) ..... 164
TABLE 5-2 MEAN WORDS PER ERROR SPOKEN BY TEACHERS IN CONFERENCES (WITH STANDARD DEVIATION AND RANGE) ..... 165
TABLE 5-3 BOTH TEACHERS: MEAN NUMBERS OF ERRORS IDENTIFIED FOR DISCUSSION COMPARED TO MEAN NUMBERS OF INITIAL PROMPTS GIVEN (WITH STANDARD DEVIATIONS AND RANGE) ..... 165
TABLE 5-4 TEACHER 1: MEAN NUMBERS OF ERRORS IDENTIFIED FOR DISCUSSION COMPARED TO MEAN NUMBERS OF INITIAL PROMPTS GIVEN (WITH STANDARD DEVIATIONS AND RANGE) ..... 166
TABLE 5-5 TEACHER 2: MEAN NUMBERS OF ERRORS IDENTIFIED FOR DISCUSSION COMPARED TO MEAN NUMBERS OF INITIAL PROMPTS GIVEN (WITH STANDARD DEVIATIONS AND RANGE) ..... 166
TABLE 5-6 BOTH TEACHERS: MEAN NUMBERS OF STUDENT ERRORS IDENTIFIED FOR DISCUSSION COMPARED TO MEAN NUMBERS OF TEACHER'S CHECKS OF UNDERSTANDING IN CONFERENCES (WITH STANDARD DEVIATION AND RANGE) ..... 166
TABLE 5-7 TEACHER 1: MEAN NUMBERS OF STUDENT ERRORS IDENTIFIED FOR DISCUSSION COMPARED TO MEAN NUMBERS OF TEACHER'S CHECKS OF UNDERSTANDING IN CONFERENCES (WITH STANDARD DEVIATION AND RANGE) ..... 167
TABLE 5-8 TEACHER 2: MEAN NUMBERS OF STUDENT ERRORS IDENTIFIED FOR DISCUSSION COMPARED TO MEAN NUMBERS OF TEACHER'S CHECKS OF UNDERSTANDING IN CONFERENCES (WITH STANDARD DEVIATION AND RANGE) ..... 167
TABLE 5-9 TOTAL NUMBERS OF ERRORS MADE BY CONFERENCE GROUP PARTICIPANTS COMPARED TO NUMBER OF CLUES OFFERED BY THE TWO TEACHERS IN THE FIRST AND SECOND CONFERENCE ..... 174
TABLE 5-10 MEAN CLUES OFFERED PER ERROR BY BOTH TEACHERS FOR BOTH CONFERENCES ..... 175

TABLE 5-11 TEACHER 1'S STUDENTS: CALCULATED WORDS PER ERROR, TURNS TAKEN, AND MEAN LONGEST RUN OF WORDS SPOKEN BY STUDENTS IN 1ST AND 2ND CONFERENCES (WITH STANDARD DEVIATION AND RANGE) ..... 178
TABLE 5-12 TEACHER 2'S STUDENTS: CALCULATED WORDS PER ERROR, TURNS TAKEN, AND MEAN LONGEST RUN OF WORDS SPOKEN BY STUDENTS IN 1ST AND 2ND CONFERENCES (WITH STANDARD DEVIATION AND RANGE) ..... 178
TABLE 5-13 BOTH TEACHERS' STUDENTS: CALCULATED WORDS PER ERROR, TURNS TAKEN, AND MEAN LONGEST RUN OF WORDS SPOKEN BY STUDENTS IN 1ST AND 2ND CONFERENCES (WITH STANDARD DEVIATION AND RANGE) ..... 178
TABLE 5-14 NUMBERS OF NON-RESPONSES GIVEN BY THE TWO TEACHERS' STUDENTS IN THE CONFERENCES ..... 182
TABLE 5-15 ADAPTING THE CONFERENCE CODING SCHEME TO THE WORKSHEET ANALYSIS ..... 186
TABLE 5-16 MEAN NUMBERS OF WORDS WRITTEN ON REFLECTIVE WORKSHEETS FOR BOTH TEACHERS FOR BOTH ITERATIONS OF THE INTERVENTION (WITH STANDARD DEVIATION AND RANGE) ..... 186
TABLE 5-17 MEAN WORKSHEET QUALITY PERCENTAGE SCORES BY ITERATION AND BY TEACHER (WITH STANDARD DEVIATION AND RANGE) ..... 187
TABLE 5-18 STUDENT PARTICIPATION IN THE TWO INTERVENTIONS, AS CALCULATED BY MEAN ERRORS MADE DIVIDED BY STUDENTS' WORDS SPOKEN IN CONFERENCES OR WRITTEN ON WORKSHEETS (STANDARD DEVIATION IN BRACKETS) ..... 191
TABLE 6-1 INTERVIEWEE CHARACTERISTICS ..... 194
TABLE 7-1 BAO'S (2014) CLASSIFICATIONS OF SILENCE AND INDICATION OF LIKELIHOOD OF PRESENCE IN THE CURRENT STUDY ..... 234

FIGURE 2-1 EXAMPLE SELF-EFFICACY INSTRUMENT ITEMS ..... 60
FIGURE 2-2 TWO EXTRACTS OF SELF-EFFICACY INSTRUMENT ITEMS FROM THE LITERATURE ..... 60
FIGURE 3-1 REFLECTIVE WORKSHEET EXAMPLE ..... 90
FIGURE 4-1 TOTAL NUMBERS OF WORDS WRITTEN FOR THE PRE-, POST-, AND DELAYED POST-TESTS FOR EACH GROUP ..... 151
FIGURE 4-2 TOTAL NUMBERS OF TARGETED ERRORS MADE IN THE PRE-, POST-, AND DELAYED POST-TESTS FOR EACH GROUP ..... 151
FIGURE 4-3 TOTAL NUMBERS OF ARTICLE-ONLY ERRORS MADE IN THE PRE-, POST-, AND DELAYED POST-TESTS FOR EACH GROUP ..... 151
FIGURE 4-4 EXAMINATION OF Q-Q PLOTS FOR NORMALITY FOR RQ1A ..... 153
FIGURE 4-5 EXAMINATION OF Q-Q PLOTS FOR NORMALITY FOR RQ1B ..... 156
FIGURE 4-6 EXAMINATION OF Q-Q PLOTS FOR NORMALITY FOR RQ2 ..... 157
FIGURE 5-1 WORKSHEET EXAMPLE ..... 185

# Abbreviations

| Abbreviation | Full form | Description |
|---|---|---|
| CEFR | Common European Framework of Reference | a guideline for describing language achievement |
| CF | Corrective feedback | correction of grammatical errors |
| CSAT | College Scholastic Ability Test | the university entrance exam in Korea; 수능 [suneung]; sometimes called K-SAT |
| DA | Discourse analysis | a formal method of text analysis |
| EFL | English as a foreign language | contexts where English is not the majority language |
| ELF | English as a lingua franca | a view of English as a contact language |
| ESL | English as a second language | contexts where English is the majority language |
| FoI | Fidelity of implementation | how closely to the original plan an intervention was implemented |
| GPA | Grade point average | the sum total value of marks acquired in a term |
| IELTS | International English Language Testing System | a standardised test of English proficiency for language learners |
| L1/ L2 | First / second language | also *mother tongue /* an acquired language |
| LAA | Language analytic ability | an ability to infer language rules |
| LC | Learning Community | a grouping of students for social and academic purposes |
| LLAMA | Llama Swansea language aptitude test | a test of language aptitude |
| MLAT | Modern Language Aptitude Test | a test of language learning aptitude |
| OOPT | Oxford Online Placement Test | a test of English proficiency for language learners |
| PS-I / PS-II | Pilot study I, Pilot study II | two pilot studies conducted in preparation for the current study |
| SCT | Sociocultural theory | a theory proposed by Vygotsky |
| SLA | Second language acquisition | a field of study |
| WCF | Written corrective feedback | correction of grammatical errors that is written on student work |

# 1 Introduction

Before beginning my doctoral studies, I worked for thirteen years with English learners in South Korea; eight of these years were spent teaching a required English writing course at a university. As the course was a writing course, corrective feedback became a focus of my research. However, I soon found that this population was under-represented in the literature, which tended to focus on students who were (a) majoring in English; or (b) in mixed first-language classes in English-majority countries, such as the USA; or (c) in general language courses which were not focused specifically on developing learners' writing. Though under-represented, the Korean learners are not alone. Internationally, large numbers of students are being required to study English in their home countries where English is not a majority language. However, English-language research of English writing in these populations is less common than on learners in English-speaking countries (Ortega, 2009). Furthermore, students who are not majoring in English, such as those studying science, technology, engineering, and mathematics (STEM) subjects may differ in their view of English from other groups of students (Peacock and Ho, 2003). For example, students who are English majors might feel more interested in the subject and be more willing to engage in learning it (ibid.). Such students might also be more willing to communicate in English, as English-language teachers with experience teaching both STEM and English-majoring students have anecdotally reported. This study aims to contribute to filling this gap.

The current study examines two ways of facilitating students' use of corrective feedback which aim to improve the grammatical accuracy in the second-language English writing of South Korean university students majoring in STEM fields whose speciality is not English language, and compares these interventions with the more standard practice of providing corrective feedback with no facilitation. This introduction aims to provide the reader with an understanding of the issues as they relate to the study. Section 1.1 introduces basic concepts relating to writing instruction, feedback in writing, and feedback practices common in classrooms; a detailed discussion of these points will be provided in Chapter 2. Section 1.2 presents a brief sketch of the broader context of English in South Korea (henceforth: Korea). Section 1.3 articulates the contribution of the study; and Section 1.4 explains the structure of the remainder of the thesis.

## 1.1 Rationale

The practice of and instruction in writing have long been accepted as beneficial for first-language (L1) writers. Of the four communication skills (listening, speaking, reading, writing), writing is the skill that is acquired last, and it is a skill that can continue to require conscious effort for its development well into advanced levels of education. Writers are prone to making errors when writing in their L1 (e.g. Greasley and Cassidy, 2009), so it is unsurprising that second language (L2) writers are also prone to error. Unfortunately for both L1 and L2 writers, non-standard writing can cause readers to make negative judgments about the authority of the writer (e.g. Chesney and Su, 2010). Thus, having a strong command of standard written language may be advantageous; understanding this, second language teachers often try to improve the writing of their students. For language learners, this tends to involve learning grammar. Representations of learnt grammar are commonly understood as being stored either as implicit knowledge, sometimes described as being subconscious, or as explicit knowledge, meaning that the learner has conscious knowledge of the learnt form and may be able to verbalise this knowledge.

The understood value of implicit knowledge is that learners can apply their knowledge without consciously thinking about it, and as such, implicit knowledge can be seen to facilitate fluent language use. The role of explicit knowledge, however, is more controversial: if explicit knowledge can become implicit, or if it serves to develop implicit knowledge, then this implies that explicit knowledge can be useful, as it can be seen as a stepping stone to implicit knowledge. Modern research tends to focus on how explicit knowledge contributes to the development of implicit knowledge.

There are cognitive theories which offer explanations about how explicit knowledge can contribute to implicit knowledge. These will be briefly outlined here, and dealt with in more detail in the literature review. Schmidt's (1990; 1995) noticing hypothesis suggests that learners' noticing of grammatical features can result in greater learning. Some common grammatical forms might naturally be noticed by learners due to the frequency with which the forms appear, or due to the prominence of the forms (Skehan, 1998:48). Teachers can also encourage the noticing of specific grammatical forms by drawing learners' attention to the targeted forms, such as through instruction (Lyster and Sato, 2013). As learners use new grammatical forms, errors can occur. A common means of focusing on grammatical accuracy in writing is the use of written corrective feedback (CF), in which learners' attention is drawn to grammatical errors in their work.

The interaction hypothesis suggests that when learners produce grammatical errors, CF provides an opportunity for the learner to be aware of the error, which can then allow them to modify their language based on the CF (Gass and Mackey, 2007). Teachers can encourage learners to learn forms by providing them with opportunities to use them and receive CF. As seen from the view of skill acquisition theory, these opportunities for practice can help pave the way for explicit knowledge to become proceduralised (Anderson, 1982:374), which could aid the development of implicit knowledge (DeKeyser, 1998).

Explicit grammar learning often occurs in writing classes with the provision of written corrective feedback (CF). Written CF can be of different types, and much research has focused on the relative effectiveness of these types, but no consensus has been reached; this will be discussed in detail in Section 2.4.2. Some of the most commonly studied are *direct* and *indirect CF*. Writing the correct grammatical form above an error in students' work is called *direct written CF*; signalling with a mark that a problem exists, and leaving the work of identifying and correcting the error to the student, is called *indirect written CF*. Some studies provide evidence that direct written CF is best; others provide evidence favouring indirect written CF. Still other studies find no significant differences between the two. Both direct and indirect CF are based on the hypothesis that the learner writer will gain generalisable knowledge from the correction that can be applied to future writing.

A weakness of this practice of CF for learners, however, is that the mere provision of CF does not ensure that learners understand why the correction was necessary. A lack of understanding could prevent learners from benefitting from the CF. Some students might take the initiative to seek out explanations for CF they do not understand; they might ask questions in class or during their teacher's office hours. They might look for answers in grammar books or on the Internet. However, there are a number of reasons why some students might not do these things; for example, they might consider these grammar points (or English in general) lower in priority than other obligations. Such students might benefit from teachers facilitating their use of CF. The interventions in this study draw from concepts in sociocultural theory for facilitating CF use; these concepts and how they aim to facilitate CF use will be explained in Section 2.4.4.3.

Following this introduction to the study, the next section will provide background information about education and English in Korea, in order to allow the reader to have an appreciation of the contextual factors influencing the population of this study.

## 1.2 Background

The setting for the current study is South Korea, so a brief overview of relevant aspects of the Korean context is useful. Section 1.2.1 provides a brief overview of the historical educational context, and Section 1.2.2 provides an overview of English teaching in Korea.

### 1.2.1 The status of education in Korea

Korean society places a high value on education. Kim and Park (2006) explain that historically in Korea, the educated elite, who were civil servants, were only able to maintain their status if they had a descendant who passed the civil service examination. In this way, '[e]ducational success benefited the individual, family, and the lineage' (ibid.:290). While descendants of the educated classes in, for example, the West[1], have also arguably benefited from the generations before them, those benefits have tended not to travel back up the line to the older generations in quite the same manner; in the Korean context, the educated elite could lose their status and benefits if the tradition was not carried on by the younger generation. In the 17th century, a Dutch merchant offered the earliest known first-hand Western account of Korea. Of education in Korea, he wrote:

> The nobility and well-to-do people give their children a good education. They take tutors in their service to teach them how to read and write, skills to which this nation is much inclined. This is done gently and with good manners. The children are constantly being told about the many wise sages of the past and how these obtained rank and honour. They mostly sit day and night and read.
>
> (Hamel, 1994:64)

The emphasis on education is also credited to Korea's roots in Confucian philosophy. Kwon, Lee, and Shin (2017:61) explain that

> Korea's traditional respect for knowledge and deep belief in continuous, lifelong human development whose emphasis on learning derives largely from the age-old Confucian belief that man is perfectible through education and that only the most learned should govern the country and society. For more than a millennium, major positions of power were allocated by civil service examinations and thus success in the examinations determined a family's fame and fortune.

However, Park (2011) warns researchers against the danger of painting Confucian heritage cultures (CHC) with one broad Confucian brush. Park points out that the ways in which Confucianism was taken up by various Asian cultures are as heterogeneous as the cultures themselves. To help researchers investigating CHC contexts avoid the problem of conflating concepts, Park distinguishes between *thin CHC discourse*, which is 'a socially constructed unity, an evolving category, hence

[1] The labels *East* and *West* are commonly used discussions such as this (e.g. Park, 2011; Williams, 2017), with *East* generally referring to countries in Asia, particularly East Asia, and *West* tending to refer to European countries generally and, in this context, particularly to Kachru's (1988) so-called inner-circle countries. Thus, these terms are applied here, but with the acknowledgement that they are imprecise.

always situationally specific'; and *thick CHC discourse*, features of which must a) contain elements that are uniquely and specifically Confucian; b) *and* not be limited to a single geographical region; c) *and* be directly connected to Confucian teachings. In the current study, where Korean society is viewed as a Confucian heritage culture, discussion of CHC will apply a thin CHC discourse, while acknowledging that, though limited to the Korean context, it cannot claim to be true for all Koreans or for all aspects of the culture.

### 1.2.2 The status and teaching of English in Korea

Korea has historically been a highly monolingual society; the US Central Intelligence Agency's website lists *Korean* as the only language of Korea (CIA, 2015). The apparent homogeneity of the society might contribute to the tendency of Koreans to behave very shyly around non-Koreans, so much so that the Korean government invested in televised public service advertisements which aimed to boost the confidence of the Korean public in communicating with foreigners (KTO, 2012). Shyness aside, there is a great concern to learn English in Korea, which has even been characterised as an 'obsession' (Seth, 2002).

Initially, Korea's preoccupation with English might not seem surprising; after all, English is widely recognised as the language of choice for international communication, education, leisure, and business for countries all over the world, including Korea. However, most Koreans do not need English for their daily lives, so why is there such a concern with studying it? The full answer to that question is beyond the scope of this thesis (Song, 2011, offers an excellent discussion); but one obvious reason for the concern with learning English is that it is a gateway subject in Korea. English proficiency opens doors to higher education and job opportunities, even where it is not an essential requirement for operating in the target context; that is, English skills are used as a tool for culling course and job applicants. For example, *SamSung Korea*'s recruitment page lists English proficiency as a basic requirement for *all* applications for professional positions, including internships (SamSung, n.d.). One reason for this gatekeeping function is that Korea has an oversupply of highly educated workers; having English requirements is one way of limiting applications. The OECD reports that 66% of the Korean population aged 25-34 have completed tertiary education; for comparison, the OECD average for the same age group is 39% (OECD, 2014a). With such intense focus on English, one might expect high levels of results; however, one ranking of countries by English ability places Korea only 30$^{th}$ out of 80 countries, in the category of *Moderate* (EF EPI, 2017). Koreans are not alone in this; even research in Hong Kong, a country with arguably stronger ties to English due to its British colonial history, report that students are entering higher education without the English L2

communication and writing skills needed for academic success (Evans and Morrison, 2011; Bruce and Hamp-Lyons, 2015).

Since English is necessary for accessing many opportunities in Korea, English instruction is in high demand. In response, many regions in the country begin compulsory English courses in elementary school and continue them through secondary school (SMOE, 2011). Unsurprisingly, compulsory English courses often continue into higher education. Highly ranked universities, such as Seoul National University (SNU) and SungKyunKwan University (SKKU) have foreign language requirements for graduation, such as required classes and required scores on standardised tests (SNU, n.d.; SKKU, n.d.). One reason for these requirements is that, although there is an English language section on the national university entrance examination, the College Scholastic Ability Test (CSAT; described below), this does not test communicative use of English (KICE, 2015). Once in university, however, students, even those not majoring in English, may be required to write reports, essays, and documents, and read and carry out research in English. Understandably, universities feel a need to cultivate the writing skills of students, and one way of doing this is by requiring students to take English writing courses as part of their higher education curriculum.

The CSAT is a national examination administered at schools across the country once per year, in November. CSAT scores largely dictate the level of university that students attend. Though recently some universities have been experimenting with considering additional factors for admissions, such as superior high school performance or essay-writing skill (e.g. Korea University, 2015), a sufficient CSAT score is still required in order to attend the institution (Kwon, Lee and Shin, 2017). On the day of the CSAT each year, everything in Korea comes to a sudden and drastic halt:

> [On the day of the CSAT,] the government rescheduled working hours, airplanes and military jets were grounded, and live-fire military drills were suspended near test sites during listening tests. Government offices and companies advised their employees to arrive at work an hour later at 10 a.m. and the stock market opened an hour later to help ease morning rush hour traffic so students could arrive at test sites in time. Construction work was also halted during Korean and English listening test hours (Lee, 2011).

In order to succeed on this all-important test, high scores in all subjects, including English, are needed. For this reason, so-called *cram schools* are popular ways of helping students maximise their learning. Cram schools are independently run, for-profit academies, which offer classes that students attend after their regular school day has ended. These schools specialise in specific subjects, notably English, mathematics, and science. The Korean government reports that 68.6% of all students attend cram schools, and that their families spend about 350,000 South Korean *won*

annually per child, or approximately 10% of their take-home income, on cram school fees, the largest share of which goes towards English classes (KOSTAT, 2015a; KOSTAT, 2015b). Cram schools are highly valued by parents; families report taking what might be considered drastic measures, such as taking on great amounts of debt or moving to different neighbourhoods, not necessarily so that their children can be in better school districts, but to be near better cram schools (Strother, 2012). Those who can afford it invest in private lessons in their homes (Kang, 2010). The 'obsession' with English is not limited to education; children and adults alike sometimes elect to undergo lingual frenectomy, a surgery performed on the tongue, believed to be a way of improving English pronunciation (Park, 2016).

The Korean education system is well known for producing high scores in maths and sciences. Since 2000, for example, Korean students have consistently scored at the top of the global Programme for International Student Assessment (PISA) standardised tests (OECD, 2014b). These strong results, however, come at high personal cost to the students and their families. High school students in some regions can stay at school until 11pm or later, and, thus, are served both lunch and dinner at their schools (Han, 2005). The expense of cram schools can put much stress on families, as described above. Students experience long hours, stress, and lack of sleep (Chakrabarti, 2013), which may contribute to the disturbingly high rate of juvenile suicide in the country (OECD, 2013).

Young students themselves have expressed to me their dread of moving from the relative freedom of elementary school to middle school, for it is then that their CSAT preparation becomes serious. In order to make time for study and extra cram school classes, students are often forced to give up music, sport, dance, and other hobbies they had previously enjoyed (OECD, 2014b). Students who fail to obtain the score needed to gain entrance to the university of their choice can choose to study for another year, independently or with the aid of cram schools and private tutors, with the hopes of achieving a higher score the following year (e.g. S.C.S., 2013). There is no limit on the number of attempts students can make; that in a given year approximately 20% of the test-takers are resitting the exam is a testament to the high level of importance students attach to the score (Kwon, Lee and Shin, 2017).

Some school teachers in Korea have expressed frustration with the exam-centric education system, saying that they feel restricted in what they can teach because of the CSAT. An experienced high school mathematics teacher explained in an interview in Ro (2017:34):

> Suneung [CSAT] is perhaps the barrier of math education in Korea, the barrier that I also must overcome. Students' only goals are getting good scores in Suneung so they can go to

> college. Math is only used as a tool for that, and they don't really care about things like mathematical thinking . . . . Students want teachers only to provide test prep, solving many, many math problems so they can get used to the type of problems in Suneung. They keep pushing me on this. But this is not a math class that I want to teach. Instead, I think mathematical thinking should come ahead [of test prep] and once it is developed, students can deal with several types of questions without any trouble.

Like mathematics, English is one of the test sections on the CSAT, so it could be expected that similar pressures might apply in English classes.

The chapter thus far has provided some context about the importance of English in Korea and of the educational context from which students in the current study had just gone through. With this background, the next section brings together the information provided about CF in Section 1.1 with information about English education in Korea in Section 1.2 to identify the contribution this study makes to the literature.

## 1.3 Contribution of the current study

Acknowledging the reality of the students in the educational context described, this study draws on the theoretical arguments summarised in Section 1.1 to explore how facilitated use of CF can encourage the development of explicit knowledge, as evidenced by reduction of errors on new pieces of writing in a population of learners who (a) are not studying English in a context where they are surrounded by the language; (b) are taking a required English writing course in university; and (c) are not English majors. Furthermore, the learners comprising this population may be experiencing mental exhaustion from their CSAT preparation years, which could reasonably be assumed to explain why, though students *acknowledge* their need of English, they seem to struggle to take charge of their learning, thus may benefit from in-class facilitation to help them develop their explicit knowledge, enabling them to apply that knowledge in their writing. It was hypothesised that learners for whom the teacher was facilitating the use of CF would develop their explicit knowledge and that this would be reflected in gains in grammatical accuracy. It was further hypothesised that these learners would demonstrate greater improvement than those not being prompted to reflect on their writing. In brief, these are the contributions this study makes to the field:

- The study links the work on development of L2 grammatical features in the contexts of explicit L2 grammar teaching with the work on corrective feedback on L2 writing, thereby contributing to the call for more L2 writing – SLA interface work to be carried out (Ortega, 2012). While the study draws from traditional SLA interests, such as investigating learners' attention to form and quantifying increases in grammatical

accuracy, it also incorporates elements from L2 writing research, such as implementing a process-writing protocol and applying a writing task over which learners have some control, such as (limited) choice of topic and an invitation to develop the topic according to personal views, which could help learners see the task as '[valuable], intrinsically interesting, and engendering a sense of ownership' (Ortega, 2012:412).

- Thus far 'little research has explored the factors that may moderate the progress a learner makes [as a result of written corrective feedback]' (Bitchener, 2016:128). CF research in SLA has tended to focus on which type of CF is the most effective rather than on how teachers can facilitate students' understanding of CF (Storch, 2010; Ferris, 2010; Brown, 2016). This study investigates different ways of facilitating feedback use rather than different *types* of feedback.
- The study uses mixed methods, combining both quantitative and qualitative analyses. Many studies of L2 writing investigate the effect of feedback techniques, but do not inquire about the experiences of the teachers and students in the study. Others investigate student and teacher reactions to feedback, but without seeking to show how effective the feedback treatment is. An understanding of both is important because, regardless of how effective a form of feedback might be, student or teacher reluctance to use it could indicate underlying problems that need to be resolved.
- The study addresses Bitchener and Storch's (2016:129) recent call for 'further product-focused research [that is] guided by questions that take into account the whole individual, the whole learner and his/her learning context'.
    - The study focuses on the needs of a category of learners that is growing worldwide: those whose primary interests lie in areas other than L2 English, but who are nonetheless taking compulsory English L2 courses in university. Research has often looked at CF for students in English-as-a-second-language (ESL) contexts, where the dominant language outside the classroom is English (Ortega, 2009), or at CF for English majors in English-as-a-foreign-language (EFL) contexts, where English is not the dominant language. The current study focuses on students who a) are not English majors, and b) are studying in an EFL context, like those where most English language instruction takes place.
    - Having a range of L1 background and prior educational experiences in a study might cloud results of studies (Cumming and Riazi, 2000). The current study focuses on a group who share an L1 and who report having similar prior educational experiences.

- The study has ecological validity, meeting the call for research which relates to learners in real classrooms (Bruton, 2009). Specifically:
    - It is longitudinal. As will be seen in the literature review, much of the work on feedback has been done in brief periods, which does not allow for a clear view of what happens in classrooms over the course of a term.
    - It investigates ways of facilitating feedback use that are possible in typical classroom settings with large numbers of students rather than being limited to laboratory settings or specialised learning environments with small numbers of learners.
    - It focuses on a limited range of grammatical structures rather than on all error types or on just one error type, both of which are common in studies on corrective feedback. This limited-focus approach may be more likely to represent what happens in classrooms such as those of the current study (Crosthwaite, in press).
    - Rather than measuring improvements in revised drafts of the same piece of writing, it examines new writing for post-feedback accuracy. Moreover, the assigned writing tasks are authentic pedagogical tasks which the students would be completing regardless of the study (Bruton, 2009).

## 1.4 Organisation of thesis

In order to understand the study and its design, an understanding of the population for whom it was designed is necessary; this was provided in this introductory chapter, including the context, background, and rationale for the study. Next, the literature review, **Chapter 2**, will provide the foundation for the study through a discussion of the existing literature. It will examine literature on L2 writing, grammar instruction, corrective feedback, and the theories that aid in understanding language development.

Having identified the gaps in knowledge about feedback in L2 writing in the literature review, **Chapter 3**, Methodology, will present how the design of this study aimed to fill those gaps for the learning context described in the Introduction. The chapter will begin by providing the aims of research and the rationale for the study design. Data collection instruments will be described, along with detailed information about the teacher and student participants. This chapter will also discuss the ethical aspects of the study and describe the procedures planned for analysing the data coming from the instruments described in this chapter.

The next three chapters will present the results. **Chapter 4** will answer those research questions (RQs) which can be investigated through quantitative statistical analysis (RQs 1-4), seeking to determine to what extent any claims can be made about the effectiveness of the interventions. **Chapter 5** will detail the analyses of post-hoc RQs 6-9, answering questions surrounding the issue of fidelity of implementation; and will explore teachers' and students' communication other during the interventions. **Chapter 6** will provide the results of the analysis of the interviews. While Chapters 4 and 5 condense data, looking for generalisations and similarities, Chapter 6 will allow for a more nuanced view of the participants as individuals. Even though the post-hoc RQs 6-9 were not part of the original plan, they are reported before the planned interview analysis (RQ 5) because gaining an understanding of the intervention through Chapters 4 and 5 will allow a better understanding of the participants' reporting of it in Chapter 6.

The three results chapters will provide information needed to understand the answers to the research questions, to gauge whether the goals of the study explained in Chapter 1 were met, and to judge how well the gaps identified in the literature discussion from Chapter 2 were filled. These issues will be discussed in more depth in **Chapter 7**, which will discuss the findings, linking them to prior research and providing the reader with a deeper understanding of how the results presented in Chapters 4, 5, and 6 contribute to knowledge in the field. Pedagogical implications of these results will also be discussed. Some of the limitations of the study will also be discussed and possible directions for future research will be considered. The chapter will close with a summary of the study, highlighting the contribution the findings make to the wider body of knowledge.

# 2 Literature review

## 2.1 Introduction

The literature review provides an overview of literature related to the topic of CF, including theoretical and empirical work relating to second language writing and second language acquisition. The chapter begins with a summary of how writing can be beneficial to L1 and L2 learners (Section 2.2). Section 2.3 provides a theoretical discussion of L2 grammar instruction, particularly as it relates to learners in instructed settings, along with a review of empirical work. Section 2.4 focuses on corrective feedback and is divided into three parts. Section 2.4.1 introduces types of corrective feedback (CF). A discussion of each type of CF is provided, including relevant empirical work. The next two sections consider factors which can influence the effectiveness of CF: Section 2.4.3 focuses on learner-dependent factors: learner context and experiences, personal preferences, self-efficacy, and language analytic ability. Section 2.4.4 focuses on factors in the control of the teacher: clarity of the CF that has been provided and whether the teacher confirms learner understanding and facilitates learner use of the CF.

## 2.2 Writing and its relevance to L2 learners

Language learners write for a number of reasons, which Manchón (2011) situates in two main strands: a) writing for the purpose of learning how to organise ideas and express them in writing, i.e. *learning to write*; and b) using writing as a tool for learning other things, i.e. *writing to learn*. This second classification can be further divided by what is being learnt: content knowledge or knowledge of the target language being used for writing. The current study is situated within a *writing to learn (language)* domain; this will be explained in Section 3.6.1.1. As L1 students and teachers are well aware, it takes much time and effort to develop the skills necessary to become a good writer, and even in L1, learning to write follows mastery of other language skills: listening, speaking, and reading. Unsurprisingly, then, the process of learning to write can be challenging. These difficulties are illustrated by quotations from Negretti's (2012) small sample (n=17) of students in a composition course in a US university (English L1 and L2 speakers, ages 17 – 55 and of various ethnicities). The participants kept journals charting their writing experiences, and reported feeling confused ('I don't understand what I need to do' p.151), frustrated ('I felt like I was going in circles' p.160), and uncertain ('I don't know how to [do what the teacher suggests]' p.160).

Despite the challenges of becoming a good writer, it seems that strong writing (i.e. learning to write) can result in academic rewards. Johnson, Tuskenis, Howell, and Jaroszewski (2011) implemented a

critical thinking and writing course for their English L1 undergraduate psychology students, in a university, presumably in the US. The participants, mostly female, self-selected by signing up for the course. During the course, the authors collected writing samples from two groups of students, course takers (n=109) and non-course takers (n=87), and asked faculty who were not a part of the study to mark the students' writing; these faculty were not informed which students had taken the thinking and writing course. The students who had taken the course received significantly higher scores on their work ($p$=.001) with a medium effect size ($d$=.54). Notably, a positive effect of this course was also visible in other courses the participants were taking: of four classes examined, the students taking the thinking and writing course before or concurrently with these four classes received higher marks than those who had not taken the thinking and writing course, and for two of those classes, the effect was significant (p=.002; p=.008). There is reason to question the use of final grades in other courses to judge the effect of the writing course; for example, it is possible that the participants were more highly motivated in those classes, or that these classes closely related to the students' majors. Importantly, though, the study suggests that writing skills may be developed through instruction, and that this development may result in academic benefits. Positive effects of writing instruction have also been highlighted in other research, such as the meta-analysis of effects of writing and writing instruction (though not focused on university students) by Graham and Hebert (2011).

Having acknowledged advantages of learning to write well, the focus now turns to writing and the development of language skills. Writing development is not synonymous with language development, and it is important to recognise that writing development 'cannot and should not be reduced to a few linguistic variables' (Norris and Manchón, 2012:225), as the complexities of writing have been well documented in research. As discussed in Section 1.1, however, errors in writing can cause readers to draw negative conclusions about the writer, so increasing the linguistic accuracy of texts can be beneficial for writers. Furthermore, as discussed in Section 1.2.2, there are sometimes institutional and social demands for linguistic accuracy. Nonetheless, it is acknowledged that linguistic accuracy is only one part of the complex task of writing and of language development.

Before delving into the theoretical discussion of language development (Section 2.3.1), it is useful to consider what is unique about the mode of writing that could make it useful to language learners. A key advantage of writing is that written work remains accessible to the learner, allowing for rereading, reflection, and revisions. Focusing on French L1 ESL learners in Quebec (n=23; judged by researcher to be of intermediate and advanced proficiency), Cumming (1990) used a think-aloud

protocol to try to understand the thought processes occurring during the L2 (English) writing process. Writers were found to focus most on their phrasing, their lexical choices, and their syntax; clearly, these are areas that would be more difficult to attend to during fluid speech, and the study suggests that L2 writers might be willing to take advantage of these opportunities to consider their work. However, it might be argued that the think-aloud protocol itself influenced the writers' language-focused behaviour; and perhaps the learners believed this would please the researcher. Kang and Pyun (2013) used observation techniques, arguably a less intrusive manner of learning about L2 writer behaviour. With only two participants, an L2 learner and a heritage language learner, the study is small. For the L2 writer, during the hour he was given to produce a short, informal text using unlimited resources, he was observed making use of online dictionaries and corpora and using instant messaging services. The researchers then used stimulated recall to ask questions about their observations, learning that the participant used the resources available to him to improve his writing with synonyms and antonyms, to check spellings, to compare his grammar against that used in example sentences, and to confirm that expressions were correct. If, as indicated in Kang & Pyun (2013) and above in Cumming (1990), learners are attending to linguistic details of the language, it seems possible that this could contribute to their language development.

## 2.3 Second language development

This section focuses on second language learning and development. There are different terms used to talk about language learning – *L2 learning, L2 acquisition, L2 development*. There is still much inconsistency, and all three can be seen used interchangeably in research. Bitchener and Storch (2016) point out that 'L2 learning and L2 acquisition are often used when referring more broadly to the overall process of learning or acquiring an L2…[while] L2 development is more frequently used to refer to the processes and stages involved in developing knowledge of the L2' (Bitchener and Storch, 2016:113). While there are numerous aspects of an L2 that can be developed (e.g. vocabulary, pronunciation), reference to L2 development in this thesis refers to the development of L2 learners' linguistic system, i.e. their interlanguage, so that their grammatical output becomes closer to the target language norm (Selinker, 1972). Section 2.3.1 will provide a discussion of inter-related theories from which language learning and development can be understood. Section 2.3.2 offers a discussion on the provision of grammar instruction for language learning development.

### 2.3.1 Theoretical discussion of second language development

This section will discuss theoretical frameworks for understanding second language acquisition and how they accommodate a possible contribution of explicit instruction to L2 development. It seems that L2 development is too complex to be explained by any single theory; indeed, theories can complement each other, furthering the understanding of language learning (Gass and Mackey, 2007). The following discussion draws on two theories from cognitive psychology, and one which incorporates the cognitive psychology theories into one that is specific to SLA. First, the discussion will explore how cognitive skill acquisition frameworks contribute to the understanding of L2 development, covering topics such as *practice* and how this relates to the development of knowledge. Next, human information-processing models of cognitive development are discussed, focusing on how the theory views limitations on learners' ability to take in new information. Lastly a discussion of the interaction hypothesis is provided, explaining what it is and how it is understood to aid L2 development. The interaction hypothesis is discussed separately from the other two theories because even though components of the theory are often associated with cognitive frameworks, its uniqueness as a theory *for* SLA warrants its own discussion.

#### *2.3.1.1 Skill acquisition*

One way of understanding language learning is through the lens of skill acquisition. As with learning to drive or to play a musical instrument, language learning can be seen as the development of skills. Anderson (1982:374), writing about cognitive skills generally rather than about language learning specifically, calls skill learning 'the knowledge compilation process that results in the translation from a declarative base for a skill to a procedural base for the skill'. Anderson's model of learning is called *adaptive control of thought* (ACT). Relating this to language learning, it is 'a gradual transition from effortful use to more automatic use of the target language' (Lyster and Sato, 2013:71). Note that when discussing cognitive skill acquisition, though research commonly refers to *phases* or *stages*, e.g. declarative *phase* or declarative *stage* (de Bot, 1996; Bitchener and Storch, 2016), this seems to be merely a tool of convenience; Fitts (1964:261), to whose model of skill development Anderson relates his own model of cognitive skill development, emphasises that

> Skill learning is primarily a continuous process, even though the fine grain structure of the performance itself may involve discrete operations. Thus it is misleading to assume distinct stages in skill learning. Instead, we should think of gradual shifts in the factor structure of the skills, or in the nature of the processes… employed, as learning progresses.

Anderson agrees, but continues to use the terms *stage* and *phase* (1983). Following this, I will also refer to *stages*, but will view these as having indistinct boundaries. This will also serve to underline

that development (L2 or otherwise) does not end once the learner reaches the advanced stage, (Fitts, 1964; Anderson, 1983), and that progress from one 'stage' to another is not a leap, in the words of Fitts, above, but a gradual shift; and therefore, overlap can occur (Kim, Ritter and Koubek, 2013).

There seems to be general agreement within this conceptual framework that cognitive (or motor) skills are developed in three main stages: a) declarative knowledge stage; b) procedural knowledge stage; and, with enough work, (a) and (b) can lead to c) automatic stage. 'There is a fundamental distinction between declarative knowledge, which refers to facts we know, and procedural knowledge, which refers to skills we know how to perform' (Anderson 1983:7). While declarative knowledge can be, though is not necessarily, acquired rather quickly, 'procedural learning occurs only in executing a skill; one learns by doing' (p348). Some researchers refer to the state of having worked through all of the steps of the procedural stage as *proceduralisation;* after proceduralisation, then, is automatisation – though they acknowledge that automatisation is not a final destination, but itself a process (DeKeyser, 2007b; de Jong and Perfetti, 2011; Bitchener and Storch, 2016). This suggests that there are levels of automatisation, which learners can continue to work through. Cohen, Servan-Schreiber, and McClelland (1992:242) make clear their view of automatisation as a process, describing automaticity as being on a continuum and 'develop[ing] gradually with practice'. They identify core components of automaticity, and these components demonstrate processes in development. For example, they explain that automaticity manifests 'an *increase* in speed of performance with practice' and '*diminishing* requirements for attention with practice' (p. 241; emphasis mine).

While practice and instruction often go hand-in-hand in classrooms, the effects of practice as separate from instruction were investigated in de Jong and Perfetti's (2011) study, focusing on proceduralisation with a small group of English L2 learners (n=24; mostly male; mean age 25) in three conditions. After preparing some notes, they audio-recorded themselves speaking. The repetition condition (n=10) gave the same speech three times; the no-repetition condition (n=9) gave three speeches on different topics, and the control condition (n=5) did nothing for this part of the study. Post hoc tests indicated that the repetition condition produced significantly more fluent speech, measured by mean pause length $t(9)= 3.647$, $p=.005$, and phonation/time ratio $t(9)=2.932$, $p=.017$. The improvement persisted on the four-week delayed post-test and with new topics. The authors hypothesise that the students 'may have been able to form new production rules and strengthen them by repeated use' (de Jong and Perfetti, 2011:561).

L2 speaking and writing are both skills; if speakers can demonstrate improvement in spoken fluency, as in de Jong and Perfetti (2011), practice with grammatical structures in writing might also produce improvements in that domain, as learners use grammatical forms and revise their work. Writers can also enjoy the benefit of permanence; unlike speech, written work remains available for review and revision (Section 2.2). DeKeyser (2007a:1) defines *practice* as 'specific activities in the second language, engaged in systematically, deliberately, with the goal of developing knowledge of and skills in the second language'. If language learners are to improve their skills, progressing from having the knowledge in declarative form to proceduralising and eventually automatising it, then they will need to practise using the language. Particularly in EFL contexts, where opportunity to use the language outside of the classroom can be limited, practice will need to happen in the classroom.

#### *2.3.1.2 Human information processing*

Another cognitive theoretical framework from the field of psychology is human information processing, credited to Miller (1956), who proposed that there were limits on the human capacity for processing information. Craik and Lockhart (1972) criticise Miller's argument about this limitation, arguing that Miller does not make it clear whether the limit in question is on processing capacity or on memory capacity. Craik and Lockhart argue that the limitation is on processing, and that memory is affected in a secondary way by the processing. In what is now referred to as *depth of processing,* they propose that differing levels of processing result in differing memory capacity. Craik and Lockhart propose two types of information processing: the first, Type-I, is the act of keeping information in consciousness by repetition; the second, Type-II, involves deeper analysis of the information. They propose that when attempting to remember information, time invested yields results when Type II processing is taking place; however, the increasing investment of time for Type I processing will not, they hypothesise, result in improved memorising of the information.

McLaughlin, Rossman, and McLeod (1983), applying human information processing theory to language learning, incorporate elements from skill acquisition theory and propose an information processing model for SLA. Their information-processing model of adult learning of cognitive skills proposes that '[s]ince humans are limited-capacity processors, [a task such as communication] requires the integration of a number of different skills, each of which has been practiced and routinized…(i.e. become automatic)' (p. 136ff) and that humans are limited in two main domains. The first is *attention*. This notion is described by James (1890:404), who writes that attention 'implies withdrawal from some things in order to deal effectively with others'. The second source of limitation is *information processing ability,* for which they draw from Craik and Lockhart's (1972)

hypothesis. McLaughlin et al. (1983) argue that if a learner has low levels of knowledge of, or experience with the content being learnt, this can influence how well or how quickly the input is processed; in contrast, learners with higher levels of knowledge of or experience with the content being learnt might be able to process the same input better and faster. These two domains together – attention and processing ability – impact on how the individual handles the incoming information. McLaughlin et al. (1983:138) summarise this:

> The more attention required, the more resources are consumed and the slower the processing. Greater practice, rehearsal, or familiarity with the material allows information to be handled more routinely without as much cost in attentional monitoring as when there is less practice, rehearsal, or familiarity with the material.

Leow (2015:204) expounds on McLaughlin et al.'s model, writing that depth of processing is 'the relative amount of cognitive effort, level of analysis, and elaboration of intake, together with the usage of prior knowledge, hypothesis testing, and rule formation employed in decoding and encoding the same grammatical or lexical item in the input'. He explains that in vocabulary learning, for example, 'shallow processing' might be when learners process how a word looks or sounds; an example of 'deep processing', then, might be when learners attempt to link the word to others in their knowledge. The cognitive effort associated with deep processing is understood to lead to greater learning benefits, as proposed by Craik and Lockhart (1972).

However, there is empirical evidence which suggests that there is a limit on how deep *deeper processing* can be and still be useful for learning. Baralt (2013) investigated effects of deeper processing on L2 development with undergraduates in intermediate Spanish L2 classes at a US institution (n=84; mostly English L1; ages 18-23). Participants participated in one of two modes: either face-to-face or via computers. Both of these groups were further split into groups receiving either simple (shallow processing) or complex (deep processing) tasks. There was also a control group, who received no treatment. Treatment group participants retold a story using picture prompts in one of two modes: the face-to-face group did this orally, and the computer group typed their retelling of the story in response to prompts shown on the screen. Feedback was provided on the targeted linguistic form in full-sentence recasts, either orally or on the screen (according to group). The complex subgroups within both modes were asked to reflect on the intentions of the characters at points throughout the story, while the simple subgroup were asked only to tell the story. All participants were tested using the same method as the treatment (but without feedback), and though they had experienced the treatment in one mode, they were asked to do the test in both the face to face and computer modes; participants also took a 15-question multiple choice test. There were three versions each communicative test and multiple choice test (pre-, post1-, and

post2-tests). Three 3 × 2 × 2 repeated-measures ANOVAs were performed:

| | |
|---|---|
| Face to face retelling | • interaction between complexity and modality, F(1, 66) = 13.20, p = .001, partial η2 = .17, power = .95<br>• statistical triple interaction for time, complexity, and modality, F(1.99, 131.18) = 7.28, p = .001, partial $\eta^2$ = .10, power = .93 |
| Computer retelling | • interaction between complexity and modality F(1, 66) = 12.13, p = .001, partial η2 = .16, power = .93<br>• statistical triple interaction for time, complexity, and modality, F(1.84, 121.5) = 7.97, p = .001, partial $\eta^2$ = .11, power = .94 |
| Multiple choice test | • interaction between complexity and modality, F(1, 66) = 12.96, p = .001, partial η2 = .16, power = .94<br>• statistical triple interaction for time, complexity, and modality, F(1.69, 111.6) = 9.93, p < .001, partial $\eta^2$ = .13, power = .97 |

The computer group with the simple task treatment received the highest score on the face-to-face mode on both post-tests. The face to face group with the complex task treatment was second.

The computer group with the simple task treatment received the highest score on the computer mode on both post-tests. The face to face group with the complex task treatment was second.

The computer group with the simple task treatment received the 2nd highest score on multiple choice post-test1, and scored the highest on post-test 2. The face-to-face group with the complex task was the highest at post-test one and second on post-test two.

It was unexpected that the simple task for the computer group garnered the highest scores, generally outperforming the complex computer group and both face-to-face groups. It seems possible that the written mode allowed learners longer processing time, which they put to use: qualitative follow-up analyses of the screen-recorded videos reveals that participants were creating their own deeper processing by scrolling to revisit the on-screen feedback on a given response before answering the next question. As the researchers suggest, the complex task might have been *too* complex for the learners. This suggests that while it might be beneficial to engage learners in deeper processing, there seems to be a limit to how much complexity learners can usefully handle.

#### *2.3.1.3 Interaction hypothesis*

Another theory contributing to the understanding of language learning is the *interaction hypothesis*, variously called *input-interaction-output model* and *interaction theory* (Gass and Mackey, 2007). Developed from within the field of SLA and incorporating elements from skill acquisition and human

information processing, the interaction hypothesis 'describes the processes involved when learners encounter input, are involved in interaction, and receive feedback and produce output' (ibid.:176), leading to the development of their interlanguage and its gradual approximation to the standard language. Interaction might commonly be associated with oral communication, given its roots in research of L2 learner conversation (Long, 1981; 1996). However, the concept has developed since its start, as the alternative names seen at the beginning of this section indicate (model to hypothesis to theory (Jordan, 2004:220)). Given these beginnings in the oral tradition, it is unsurprising that a strong connection with speech remains, but there is evidence that the theory is expanding to encompass input and output in written form as well as spoken form: in addition to being applied to research on written corrective feedback (e.g. Bitchener & Storch, 2016), in her volume on the topic, Gass (2018:n.p.) herself includes discussion on written forms as well as spoken ones.

Within the interaction hypothesis, Gass (1997, 2018) explains five stages learners go through in the process of learning new information; Bitchener and Storch (2016:n.p.) base their model of a single episode of written correction upon these stages (discussed in Section 3.6.6). The first stage is what learners' context offers them – input. Gass and Mackey (2007) explain that input is a source of positive evidence, or 'information about what is possible in a language' (p. 177). Input is 'the language to which a learner is exposed either orally or visually...' (Gass, n.p.). However, the mere fact that input is available does not necessarily result in it being 'taken in' by the learner. Thus, Gass (2018) specifies two kinds of input, which are the first two stages of the framework: (1) ***apperceived*** **input** and (2) ***comprehended*** **input**. Apperception is defined as 'the process of understanding by which newly observed qualities of an object are initially related to past experiences' (ibid.:n.p.). It is influenced by knowledge the learner already has, which causes the learner to take note of something in the input (Gass, 2018). Comprehended input involves analysis of the information as the learner strives to grasp the meaning. Comprehend*ed* input stands in contrast to the familiar comprehen*sible* input (Krashen, 1982), which Gass (2018) criticises as being speaker-centric and limited to a yes/no (i.e. comprehended or not) outcome. Comprehended input occurs on a continuum, and each stage of the continuum, even if the learner has not succeeded in grasping the full meaning, may have something of benefit for the learner (ibid.); i.e. a piece of the puzzle that may contribute to later understanding.

(3) **Intake** refers to 'the process of assimilating linguistic material' (Gass 2018:n.p.). What is taken in (i.e. the comprehended input) gets compared to the learners' existing knowledge and assimilated (ibid). After being taken in and considered against the learners' existing knowledge, the new

knowledge is (4) **integrated**, meaning it might be a) used to (dis)confirm hypotheses; b) strengthen existing knowledge; c) stored away; or d) lost (ibid). Finally, the learner produces (5) **output** (ibid). Output provides an opportunity for learners to test their hypotheses about the knowledge they are gaining about the language, and, linking this model with skill acquisition theory, through output practice, learners can work toward automaticity (Gass and Mackey, 2007). Feedback on practice can provide the learner with information about their hypotheses, allowing further hypotheses to be developed.

The *interaction* aspect of the interaction hypothesis can take place anywhere between input and output (Gass, 2018:n.p.). For example, when taking in new information (input), learners can be exposed to positive and negative evidence. While positive evidence was explained above as showing what is possible in a language, negative evidence, by contrast, shows what is unacceptable. Gass and Mackey (2007) see the negative evidence coming from interaction, and so available at multiple stages; Bitchener and Storch (2016), focusing on writing, place negative evidence as occurring at input in the form of corrections. Regardless of the stage at which it is classified, negative evidence can benefit learners by helping them become aware of possible errors in their speech; such awareness could be brought about by attempts to reach a mutual understanding, or by an explicit correction (ibid.). An example of explicit corrections offering negative evidence is negative feedback. 'Negative feedback … may be facilitative of L2 development… and essential for learning certain specifiable L1-L2 contrasts' (Long, 1996:414). That is, if the learner's output is corrected, this can lead them to modify their output (Swain and Lapkin, 1995).

The process of a) a learner offering an utterance that contains an error; b) being made aware of it through interaction; and then c) modifying their output is considered by Swain and Lapkin to be an important part of language learning. The reason for this may be drawn from human information processing, as learners are encouraged to analyse their productions, thereby resulting in greater learning. Although de Jong and Perfetti's (2011) participants, discussed in Section 2.3.1.1, did show improvement with only practice, improvement of fluency and improvement of grammatical accuracy are different skills. While practice alone might result in improved fluency, those aiming to increase grammatical accuracy might require feedback in order for practice to be useful. One of the groups in Van Beuningen, de Jong, and Kuiken's (2012) study with Dutch L1 and L2 secondary school students (n=268) was a practice-only group who did not receive feedback on the accuracy of their grammar. An ANOVA revealed that without correction, participants were unable to improve their grammatical accuracy to a significant level, unlike the groups who did receive feedback.

Researching the interaction hypothesis involves investigating 'the relationship between communication and acquisition and the mechanisms … that mediate between them' (Gass, 2005:224). One such mechanism is when the learner notices new information. Apperceived input could be seen as the gateway, where the learner either begins the process of taking in the new information, or, if the information is not detected, ignores it. Gass (2018) sees this *noticing* as an integral part of apperceived input, and calls apperceived input the cornerstone of the learning process. Noticing (Schmidt, 1990; 1995) is centring one's attention on a grammatical feature, and Schmidt hypothesises that levels of learners' noticing can correspond to levels of learning. Some researchers have argued that this is too simplistic and have proposed a different organisation of the concept. Tomlin and Villa (1994), based on a review of literature, propose a three-pronged framework: *alertness* ('readiness to deal with incoming stimuli or data'; p. 190); *orientation* (where attention is directed); and *detection* ('cognitive registration of sensory stimuli'; p. 192). Alertness could be expected to vary for each student over time during a course and even within a single class; the need for orientation could be fulfilled in a classroom environment by giving lessons, thereby orienting students to the topic to be focused on (Gass, 2018). It is the third prong, detection, which Tomlin and Villa (1994) identify as key to learning. However, Robinson (1995) criticises Tomlin and Villa's model, pointing out that as the critical component is considered to be detection, then it is much the same as Schmidt's *noticing*. Robinson (1995) instead proposes keeping the concept of noticing, but defining it as 'detection plus rehearsal in short-term memory, prior to encoding in long-term memory' (p. 296). Perhaps considering the fine distinctions of these proposed alternative views of noticing difficult to operationalise and apply in research, many researchers apply Schmidt's *noticing* to their studies (Sachs and Polio, 2007; Santos, López-Serrano and Manchón, 2010; Park, 2013).

Elaborating on the hypothesis, Schmidt (1990) explains that there are factors which can impact on a learner's ability to notice: 1) *frequency*; that is, the more times the form appears, the greater the chances of it being noticed); 2) *perceptual salience*; which Schmidt offers examples of, but does not explain in detail. Skehan (1998:48) explains *perceptual salience* as 'how prominent a form is in input. […] [T]he more a form stands out in the input stream, the more likely it is that it will be noticed'; 3) *skill level*, both the skill being targeted and the learner's ability to focus on multiple inputs at once; 4) how well the learner is able to process information for the task.

Cumming's (1990) study suggests that L2 learners can be aware that their writing might contain errors, and that they may actively try to seek out the correct forms (Section 2.2). Understandably,

though, they can only correct as much as their current knowledge will allow; logically, learners will generally at least suspect that there may be an error before they note the need for correction. Swain and Lapkin (1995) explored the noticing of English L1 adolescent French immersion learners (n=9) using a think-aloud protocol, asking the learners to verbalise what they were thinking while writing and then editing an expository text. The authors were interested in language-related episodes (LREs), which they defined as when 'a learner either spoke about a language problem he/she encountered while writing and solved it [accurately or not], or simply solved it…without having explicitly identified it as a problem' (ibid.:378). The learners in the study produced 190 LREs, and of these, around 41% occurred when learners were prompted to edit their work, rather than as a part of the initial writing process. This could indicate that learners had noticed and were trying to correct a problem. The authors write that 'on each occasion, students engaged in mental processing that may have generated linguistic knowledge that is new for the learner, or consolidated existing knowledge' (ibid.:384).

Studies have suggested that a possible disadvantage of think-aloud protocols such as those used by Swain and Lapkin (1995) could be the added cognitive load of speaking aloud (e.g. Koro-Ljungberg et al., 2013). Yet, despite this possible added cognitive load, the learners in Swain and Lapkin's (1995) study above seem to have been analysing linguistic information in their writing. It is reasonable to think that learners who are not asked to think aloud might also analyse the linguistic features of their writing. While the authors did not report how accurately the learners were able to correct their work, they do note that learners' thinking was 'sometimes faulty, leading to incorrect hypotheses and inappropriate generalizations' (p. 384). Nonetheless, the study seems to demonstrate that learners have the desire for accuracy in their writing, and suggests that expert input could support learning even more by guiding learners' noticing.

If experts are to assist learners in noticing, it could be relevant to explore what it is that learners notice on their own. Hanaoka (2007) investigated what second year university Japanese L1 English learners (n=37; intermediate and advanced proficiency) noticed when they compared their narrative-style writing samples (based on picture prompts) to 'native-speaker' models. Students made notes about differences they noticed between the two pieces of writing, and researchers found that participants spontaneously attended mostly to lexical differences between their writing and the models: 92.4% of what they noticed was lexical. Grammar accounted for only 2.3% of what was noticed (with the remaining being categorised as 'Other'). It could be argued that their choice of lexis was more constrained because of the pictures, whereas problematic grammatical forms might

have been more easily avoided. However, the finding that learners tend to focus on lexis more than grammar is a common finding in research, and will be discussed later in this section.

If grammar is to be focused on, drawing learners' attention to the targets could be useful. Gass, Svetics, and Lemelin (2003) investigated the attention of English L1 learners of Italian (n=34; 8 male; various proficiencies) in university language courses with a pre-test/treatment/post-test between-subjects study. After the pre-test, researchers used a computer programme to manipulate learners' attention in texts, in one condition underlining parts of the text and telling students to learn them, and in the other condition, not underlining anything and asking students to focus on comprehension. The underlining condition focused on three areas: Italian L2 syntax (question formation), morphosyntax (direct and indirect object movement), and lexis (five new lexical items). These treatments lasted approximately 50 minutes, followed immediately by grammaticality judgement and lexis post-tests. Wilcoxon matched-pairs signed-ranks test results showed that the learners in both conditions improved significantly from pre-test to post-test on morphosyntax (underline: $z$ = -.524, $p$<.001, $d$=1.48; no underline: z=2.631, $p$<.01, $d$.38) and lexis (underline: $z$=-3.328, $p$<.001; $d$=.97, $z$=-2.083; no underline: $z$=-2.083, $p$<.05; $d$=.49), but only learners in the underlining condition improved significantly in syntax ($z$=-3.519, $p$<.001; $d$=1.57). This suggests that while learners *noticed* morphosyntax and lexicon independently and appeared to benefit from this, for syntax, 'learning without focused attention did not occur' (ibid.:520). They explain that this seems to support the benefit of drawing learners' attention to specific language features, particularly when those features are complex. While there were few participants, and they were put into groups randomly rather than stratified by proficiency, the studies here suggest the possibility that learners might need to be guided to focus on grammatical accuracy if that is what the teacher wishes to focus on.

Examining the effects of different types of expert input on learners' revisions, Sachs and Polio (2007), in the second part of their two-part study, divided their ESL university learners of various proficiencies (n=54; mostly Korean and Japanese) into four groups (direct written corrective feedback (WCF), expert-provided reformulation, expert-provided reformulation + think-aloud, and a control group). The learners received a picture story and were asked to write the story. Two days later, they received feedback according to treatment group, and then the following day the learners rewrote the story without looking at their feedback or reformulation again. For the unit of measurement, the researchers used t-units, that is, 'one main clause plus all the subordinate clauses attached to or embedded within it' (Hunt, 1965:141). Non-parametric tests showed that the three treatment groups revised significantly more t-units ($p$<.05) than the control group, and the revisions

were also more accurate. Using two-tailed Mann-Whitney tests, each condition was tested in order of ranked result: WCF had the highest mean rank of all conditions (42.63), and it was significantly different ($p<.05$) from the reformulation group (28.14); note that the reformulation score was approximately half that of WCF. The difference between reformulation + think-aloud (26.84) and the control (15.63) was also significantly different ($p<.01$), but the difference between the two reformulation conditions was found to be not significant. Taking Schmidt's (1995) noticing hypothesis (*more noticing leads to more learning*) as a possible explanation, it is possible that reformulation masked the corrections so that the learners were less able to notice them, while the direct WCF may have stood out and aided the learners' noticing.

Hanaoka and Izumi (2012) also found positive effects for noticing with EFL learners (n=38) in Japan, but, interestingly, they report finding a more positive effect with reformulations than Sachs and Polio (2007). The task designs were similar to those in Sachs and Polio (2007): both studies used pictures to prompt story-writing, and the resulting writing samples were of comparable lengths. An important difference to the Sachs and Polio (2007) study is that the students were asked to keep a written L1 record of the problems they encountered during their writing. One week later, for the treatment, Hanaoka and Izumi (2012) gave participants two feedback texts: a reformulation of their writing and a model writing sample. Participants were not told that one of these was a reformulation; the two feedback texts were merely labelled *Model 1* and *Model 2*. Half of the participants received the reformulation first, and then the model; the other half received the model first, and then the reformulation. The participants were given seven minutes to compare their original text to the first feedback text they received (Model 1 or 2, as the texts were counterbalanced), noting in a designated space on the feedback paper in L1 any differences they saw. Then the first model texts were removed and the second ones distributed; participants were given seven minutes to repeat the process with the second model they received. The participants were again asked to keep a written L1 record of the differences they noticed. Comparing original writing and revisions, chi-square tests revealed that changes made from reformulations (98% of errors corrected) was significantly greater ($p<.001$) than from model answers (48% of errors corrected). The result is perhaps unsurprising since reformulations directly correspond to what each student has written while model answers might have less in common with what a student has actually written.

There are some important differences between the two studies. Hanaoka and Izumi's (2012) study lacked a control group and did not use direct WCF. There are some large differences in the manner of feedback presentation make the studies difficult to compare: participants in Hanaoka and Izumi's

(2012) study kept L1 records of problems they experienced while writing, and then again while examining the reformulation and model answers. Being asked to be aware of what they were uncertain about might have primed students to attend to the differences in the model and reformulations. Importantly, Sachs and Polio (2007) found more (and more accurate) corrections with their direct WCF group than for either of the reformulation conditions; this is in line with findings in other studies comparing direct correction with reformulations (e.g. Santos, López-Serrano and Manchón, 2010), and may be influenced by L2 proficiency, with lower-proficient learners in particular seeming to focus on lexis (Swain and Lapkin, 1995; Santos, López-Serrano and Manchón, 2010; Baek, Song and Lee, 2012). This might be explained by VanPatten's (2002) principles of input: that learners have a tendency to attend to meaning (lexis) before they attempt to process the form (syntax) of the input. What the studies do suggest is that learners in both ESL and EFL contexts are capable of noticing, and that teachers might be able to capitalise on this ability in classroom contexts by guiding the students to attend to specific problems through feedback.

The three theoretical frameworks discussed here provide a basis for the practice of written corrective feedback, which will be explored in Section 2.4. The discussion thus far suggests that there are stages learners go through when taking in new information, the provision of positive and negative evidence learners can be useful to learners. The research presented here suggests that learners a) do try to attend to their errors; b) that this *noticing* can promote learning; c) that learners can fail to notice grammatical issues independently; and d) that experts can influence what it is that learners notice. Thus, it may be beneficial for experts to guide students' noticing if the focus is to be grammatical accuracy. This is discussed in the next section.

### 2.3.2 Grammar instruction for language development

The previous section discussed how theories relating to noticing, practising, and deeply processing language can contribute to language development; this section focuses on using grammar instruction to further these processes. Following a brief discussion about some implications of grammar instruction in a globalised world, Section 2.3.2.1 discusses how grammar instruction is understood to contribute to language development. Section 2.3.2.2 provides a discussion of how grammar instruction is provided.

In a globalised world, one might ask: *Does standard grammar matter?* It seems that the answer to this important question depends on the perspective of the person asking it. Standard English can be 'conceived of as unmarked, stable, and uniform' (Johnston and Lange, 2006:193), but as Seidlhofer

(2017) points out, stability can be seen as a chimaera, because languages are in flux. Given this, there might be arguments against a strong focus on standard grammar, particularly from an *English as a Lingua Franca* (ELF) perspective. This perspective focuses on English as a contact language, and concerns 'English use within and across … borders, as well as focusing on the global consequences of English's use as a world language' (Galloway and Rose, 2015:xii). In Björkman (2008), in a university ELF setting in Sweden, where students and faculty use English to communicate but have various L1s, a corpus of 28 hours of recorded group work collected from six departments was analysed. Non-standard morphosyntactic features were classified into two categories: *Disturbance* when the non-standard utterance prompted clarification action from interlocutors, or *Non-Disturbance* when communication carried on without disruption. Analyses of the eleven identified types of non-standard morphosyntactic utterances revealed that only one – question formulation – caused communication breakdown. This could suggest that non-standard English usage can be unproblematic.

If non-standard use is not usually problematic, is a focus on standard grammar necessary? ELF is a relatively new perspective; though there was discussion of it in the 1990s, it was not until 2000 that interest in ELF research began to grow (Galloway and Rose, 2015:149). It might take time for these viewpoints to become accepted by society in general, and by business and academic gatekeepers, such as those in Section 1.2.2. In Greasley and Cassidy (2009), lecturers in the faculties of health and social science (n=32) at a university in England identified *poor language, grammar, and expression* as the number one thing that frustrated them about (presumably) L1 student work. These lecturers are not alone; other studies have found similar results, with some lecturers even singling out L2 learners as a source of frustration. Ferris, Brown, Liu, & Stine (2011) surveyed (n=129) and interviewed (n=23) writing instructors at two- and four-year colleges in the US. Of 44 comments made by instructors describing how they provide feedback differently to second-language writers compared to English L1 writers, 42 of the comments dealt with linguistic problems; the final two comments discussed how they provide encouragement. The study found that instructors can sometimes feel resentful towards the perceived additional burden of having ESL students in their classes; one tutor even actively encouraged ESL students to drop the course. The tutors' focus on linguistic problems in this study may suggest that tutors can find grammatical issues particularly irksome, which could result in L2 writers receiving negative assessments despite the quality of their ideas. Master (1995), focusing on article errors, warns that non-standard use of grammar could lead the reader to draw negative conclusions about the proficiency of the L2 writer; more recent research also supports this, as

Zawacki and Habib (2014) found in interviews with faculty (n=18) from sixteen disciplines in a US higher education institution.

Learners themselves may expect language teachers to focus on learners' non-standard grammar and help them improve; 83% of English learners in one study (Chinese L1; n=320) reported having an expectation that teachers correct *all* of their errors (Lee, 2004). Encouragingly, recent research suggests that teachers can moderate students' belief that the goal of English L2 study is perfection of an ideal standard or native-likeness (Rose and Galloway, in press), though a full discussion of the topic is beyond the scope of this thesis. At least for now, the desire for standard grammar appears to be a reality that many language teachers and students must deal with.

#### *2.3.2.1 How grammar instruction is useful*

Ellis (2006), criticising earlier conceptualisations of explicit grammar instruction (henceforth *grammar instruction*) as *presenting and practicing*, as in Ur (1996) and in Hedge (2000), as being too narrow, offers this definition:

> Grammar teaching involves any instructional technique that draws learners' attention to some specific grammatical form in such a way that it helps them either to understand it metalinguistically and/or process it in comprehension and/or production so that they can internalize it (2006:84).

The definition focuses on two aspects of grammar instruction: helping learners to internalise a grammatical form by a) understanding it metalinguistically; and/or b) processing it through comprehension and production). While compared to the definitions Ellis (2006) criticises for being too narrow, this definition does succeed in increased scope, but the theoretical bases of these aspects of grammar instruction merit unpacking. What does it mean to understand and internalise grammatical forms?

***Implicit vs explicit L2 knowledge***

Representations of grammar are generally understood as being stored as two kinds of knowledge: one is implicit knowledge, which has been described as *subconscious* and compared to Bialystok's (1982:183) concept of *unanalysed knowledge*, or 'the general form in which we know most things without being aware of the structure of that knowledge'. Ellis (2006:95) explains that implicit knowledge is 'procedural, is held unconsciously, and can only be verbalized if it is made explicit'. The other commonly recognized knowledge form is *explicit knowledge*. Ellis (2006:95) defines explicit knowledge as that which is 'held consciously, is learnable and verbalisable, and is typically accessed

through controlled processing when learners experience some kind of linguistic difficulty in using the L2'. Both of these definitions include an element of consciousness, but this is not an uncontested stance; for example, some researchers argue that it is not possible to truly distinguish between *conscious* and *unconscious* (e.g. McLaughlin, Rossman and McLeod, 1983).

Although these definitions employ some terms that overlap with declarative and procedural knowledge, explicit knowledge and declarative knowledge cannot be considered synonymous terms. While the terms implicit/ explicit knowledge indicate that the knowledge is less or more conscious (acknowledging the above problem with consciousness), the terms declarative/ procedural describe how the learner can use that knowledge (Ellis, 1993). While 'there is broad consensus that the acquisition of an L2 entails the development of implicit knowledge' (Ellis, 2005:143), there is disagreement about the way in which explicit knowledge contributes to language learning (Ellis, 2004). This has led to the so-called interface positions, with *interface* meaning that 'there is a connection or overlap between the two types of knowledge' (Han and Finneran, 2014:371): *strong interface, weak interface*, and *non-interface*, which are discussed briefly in turn below before considering a balanced alternative.

***Strong interface*:** Some researchers argue for a strong interface between explicit and implicit knowledge, where the former, through repetition, becomes the latter (DeKeyser, 2007a). This position draws strongly on skill acquisition theory. An example can be found in de Jong and Perfetti's (2011) study, as discussed in Section 2.3.1.1, with participants demonstrating linguistic development through practice, even without the benefit of instruction and feedback. The effect has been demonstrated in larger studies as well (White et al., 1991).

***Weak interface*:** Other researchers argue that explicit knowledge can become implicit knowledge, but only under certain conditions, such as using certain kinds of pedagogies which make the targeted forms salient and prompt cognitive processing (Norris and Ortega, 2001). An example of this type of research is Lindseth (2016), who investigated the issue with L2 university students of German (n=13; presumably US; presumably English L1). The treatment group (n=6) received 20 minutes of form-focused instruction on subject-verb (SV) inversion during each of five classes, while the control group (n=7) received nothing. Tests were 15-20 minute recorded interviews. Participants' use of SV inversion was found to be similar between groups at the pretest, but at the post-test, the treatment group correctly used SV inversion in obligatory contexts (67.2%) significantly more than the control group's 35% ($z=2.429$, $p=0.008$), and with a large effect size: $d=1.631$. Because of the

'online' nature of the test (i.e. participants not having time to consider explicit knowledge), the authors suggest that the test may draw on implicit knowledge. Though small numbers of participants mean that caution in interpretation is required, it seems possible that, as the authors suggest, explicit form-focused instruction supported the development of participants' implicit knowledge. This will be further explored in Section 2.3.2.2.

***Non-interface:*** Still other researchers view explicit and implicit knowledge as entirely separate, arguing that the one cannot become the other. Even this position, however, acknowledges a role for explicit knowledge. For example, explicit knowledge might be seen as contributing to language learning by facilitating learners' monitoring ability during output (Krashen, 1982:15). Paradis (2004:51) also sees a place for explicit knowledge, and accepts that practice can

> facilitate the acquisition of implicit competence, though not because explicit knowledge eventually becomes implicit competence. […] However, it is not the *knowledge* that leads to competence, but the *practicing* of the relevant form— without conscious awareness of what is being internalized (italics original).

Van Patten (2016) expands on the argument, arguing that results such as those seen in Lindseth (2016) cannot be considered indicative of implicit knowledge development, explaining that 'rules' provided to learners are not actually rules at all, but rather mere descriptions of rules. He argues that rules are mental representations that cannot be conveyed through instruction, but rather developed only through input. While the argument may have merit, there are two main problems with it. Firstly, it fails to suggest a useful way forward, making the position one of questionable pedagogical value, particularly to EFL settings, where target language input might be in short supply. Secondly, and possibly more importantly, it fails to address the issue of empirical evidence which does seem to support some level of interaction between explicit and implicit knowledge, as seen in the discussion above; researchers who reject the non-interface position cite studies such as these as their basis for their position (e.g. Doughty and Williams, 1998; Norris and Ortega, 2001).

**A way forward**

Despite being listed as distinct classifications, it seems there is room for some stronger and weaker views within each position, so it might be more useful to view the three positions on a continuum. It might also be useful to consider that each position may be relevant in different ways. Han and Finneran (2014:372) argue in their theoretical discussion that

> for instructed L2 learners, there is not just a singular relationship between explicit and implicit knowledge, but several co-existing relationships – a strong interface (for some linguistic elements), a weak interface (for some others), and no interface (for others) – and that each of these can be multi-faceted.

Han and Finneran (2014) argue that researchers' tendency to depend on one of these interface positions to the exclusion of the others is detrimental to furthering research in this area. The three positions each have merit, as evidenced by the discussion of each above, and the field has much more to gain by putting aside the debate on which one is 'right', and focusing on '*which aspects of grammar are susceptible to a strong interface, a weak interface, or no interface across and within second language learners*? (p. 383; emphasis original).

Two researchers representing two interface positions (weak and strong) approached the issue as a team, designing a large study to investigate, inter alia, effects of explicit knowledge. Suzuki and DeKeyser (2017) conducted their study with Japanese L2 (Chinese L1) speakers in Japan (n=100; mean age 25) who had learnt Japanese in instructed (i.e. classroom) settings. Targeting three linguistic forms, the researchers conducted tests to measure, inter alia, implicit and automatised explicit knowledge (Table 2-1), designing them to try to prevent participants from attending to grammatical forms. The researchers focused on *automatised* explicit knowledge because, as it has been in the process of becoming proceduralised, access can be rapid even if awareness is still a part of the process. Implicit knowledge was, as seen in the table below, differentiated with online (i.e. real-time) tests.

**Table 2-1 Characteristics of Suzuki and DeKeyser's (2017) tests (reproduced from Suzuki and DeKeyser, 2017:761)**

| | Implicit knowledge | | | Automatised explicit knowledge | | |
|---|---|---|---|---|---|---|
| Characteristic | Visual-world | Word-monitoring | Self-paced reading | Timed AGJT | Timed VGJT | Timed SPOT |
| Data | Fixation | RT | RT | Accuracy | Accuracy | Accuracy |
| Focus | Meaning | Meaning | Meaning | Form | Form | Form |
| Real-time | Yes | Yes | Yes | No | No | No |
| Modality | Aural | Aural | Written | Aural | Written | Written |

RT = reaction time; AGJT = auditory grammaticality judgment task; VGJT = visual grammaticality judgment task; SPOT =fill-in-the-blank task

- Visual-world: With eye-tracking to monitor attention, participants were shown four pictures on a computer screen for 5.5 seconds while an audio two-sentence story was played. Eye movements were tracked as an indication of how sensitive participants were to the grammar in the sentence, i.e. they might focus on the subject or the object picture depending on the grammar of the sentence. After this, participants answered the yes/ no comprehension question that appeared; comprehension questions were included here and elsewhere to distract participants from focusing on grammar. Participants completed 48 trials with 16 'filler' trials mixed in to redirect attention.

- Word-monitoring: Participants were shown a target word, and then they listened to a recorded sentence, pressing a button as soon as they heard the target word, providing a measure of their reaction time. They then answered a T/F (distractor) comprehension question. There were 48 trials and 48 filler items.
- Self-paced reading: Participants were asked to read a sentence as quickly as they were able and answer the comprehension question. Sentences were given one word at a time as the participant pressed a key, with the previous word disappearing as each new one appeared; thus, the participant controlled the speed of their reading. At the end of the sentence, a comprehension question appeared. This provided a measure of how quickly the sentence had been read before the question was answered. There were 48 trials and 48 filler items.
- Timed Auditory Grammatical Judgment Task: Participants were asked to decide whether a sentence they heard was grammatical or not, and to make the decision as quickly as they were able. There were 48 trials, and 16 filler items.
- Timed Written Grammatical Judgment Task: Same as above, but the sentences were written.
- Timed SPOT (Fill-in-the-Blanks) Test: Participants filled in the blank with the targeted form; there were 48 trials and 16 filler items.

Using structural equation modelling (SEM), the researchers created two models, an interface model and a non-interface model. The interface model hypothesises directional relationships between automatised explicit knowledge and implicit knowledge, while the non-interface model does not. This model comparison was done to examine the relationship between automatised explicit knowledge and implicit knowledge as well as measuring the level of contribution automatised explicit knowledge might make in the formation of implicit knowledge. That is, the non-interface model can be viewed as providing a baseline, allowing difference in implicit knowledge contribution between the two models to be measured. Although there was no significant difference found between the two models, which means definitive conclusions cannot be drawn about the role of explicit knowledge, there was a significant relationship between automatised explicit knowledge and implicit knowledge with the interface model, $r=.35$, $z=2.42$, $p=.02$. Suzuki and DeKeyser (2017), while acknowledging that the results of their study cannot be considered 'proof' of an interface, they argue that the results indicate that '[a]utomatized explicit knowledge, fostered by explicit learning mechanisms, influences the acquisition of implicit knowledge' (p. 781).

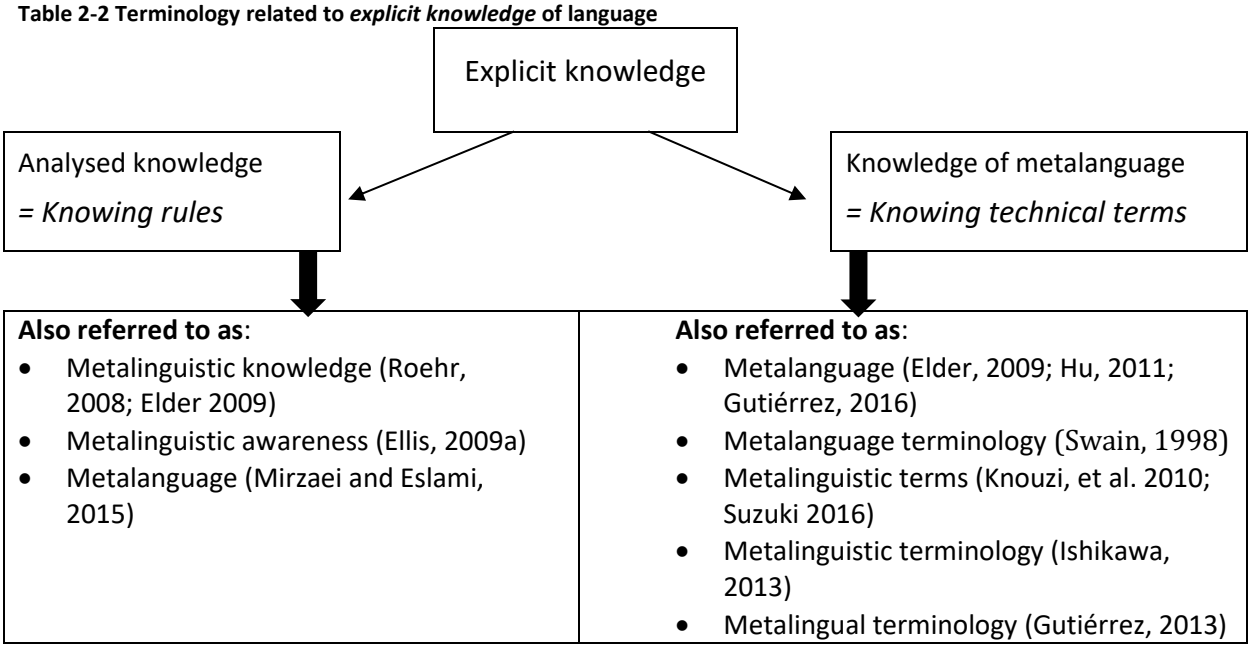
Evidence from all three interface positions and from Suzuki and DeKeyser (2017) demonstrates that explicit knowledge has a role to play. As in the discussion on comprehended input from the

discussion of interaction hypothesis in Section 2.3.1.3, it is possible that the explicit knowledge 'pieces' of the language puzzle each contribute to the whole, even if the development is neither complete nor immediate. Thus, it seems there is sufficient evidence for the positive contribution of explicit knowledge to L2 grammatical development to permit further examination of it.

**Kinds of explicit knowledge**

There are two commonly acknowledged kinds of explicit knowledge about language: analysed knowledge and knowledge of metalanguage: *analysed knowledge* is knowledge of grammatical rules, while *knowledge of metalanguage* is knowledge of technical terms used to express those rules (Ringbom, 1987:51; Han and Ellis, 1998; Ellis, 2004; Gutiérrez, 2016). Though it is considered a type of explicit knowledge, knowledge of metalanguage is not essential for the development of analysed knowledge; nor is it necessary for knowledge of metalanguage to be in place for analysed knowledge to develop and be measured (Swain, 1998; Ellis, 2004). A variety of terms are used to express this distinction, as shown in Table 2-2.

**Table 2-2 Terminology related to *explicit knowledge* of language**




Given the lack of terminological consistency in the field, it is necessary to clarify which term will be used for which concept in the current study. *Analysed knowledge*, as used in this context, seems rarely to appear independently in discussions of explicit knowledge. *Metalanguage* is a poor choice for this study, as its appearance in both columns suggests confusion. With roots in cognitive psychology, *metalinguistic awareness* is used when discussing children's language development, either L1 or L2, or both, and for bilinguals (Cummins, 1978; Bialystok, 1988; Bialystok and Barac, 2012). The argument here is not that *metalinguistic awareness* cannot be used in other contexts, but

rather that when equally appropriate options are available, as here, it is ideal to choose one less likely to cause confusion. Thus, *knowing rules* will be referred to as *metalinguistic knowledge*.

When referring to *knowledge of metalanguage*, most terms include a noun that helps clarify the meaning – *terminology* or *term* – making these equally good choices. For clarity, and for coherence with *metalinguistic knowledge*, selected above, *metalinguistic terminology* will be used in this study to refer *to knowing technical terms*. The outcome of this discussion is summarised in Table 2-3.

**Table 2-3 Explanation of terms used when discussing aspects of explicit knowledge**

| Aspect of explicit knowledge | Refers to knowledge of | Selected clarifying term |
|---|---|---|
| Analysed knowledge | *rules* | Metalinguistic knowledge |
| Knowledge of metalanguage | *technical terms* | Metalinguistic terminology |

#### *2.3.2.2 How grammar instruction is provided*

Ellis (2004) argues that explicit knowledge can be developed through grammar instruction, and argues that this can be useful on two planes: firstly, instruction can provide learners with opportunities to develop declarative knowledge of grammar features. Secondly, instruction can be used to proceduralise learners' knowledge, by providing opportunities to practice and perfect what has been learnt. This view of developing language skills through a process of developing declarative knowledge and then practising and proceduralising it is in line with Anderson's (1982) skill acquisition theory, discussed in Section 2.3.1.1. For developing explicit knowledge, a number of ways of presenting grammar features explicitly through classroom instruction have been proposed, and the classification of these tends to be centered either on *how* linguistic targets are selected for focus, or *when* instruction takes place in a lesson.

**Focusing learners on grammar**

Language learners have been said to be able to learn grammar implicitly (Krashen, 1982), meaning that attention to the object of learning is not required, leaving learners unaware when learning has occurred and thus unable to verbalise what it is that has been learnt (Ellis, 2009a). However, learners in some contexts where the language being learnt is less commonly used might struggle to get as much exposure to the language outside of the classroom as learners immersed in the context of the target language do, making reliance on implicit learning inefficient.

Explicit grammar learning is when 'rules are explained to learners, or when learners are directed to find rules by attending to forms' (Doughty, 2005:256). The increased efficiency of explicit grammar

learning over implicit learning has been demonstrated both in individual empirical studies (e.g. DeKeyser, 1994; 1995; Robinson, 1997; Klapper and Rees, 2003; Andrews, 2007; Akakura, 2012; Nazari, 2013) and in reviews and meta-analyses of studies (Norris and Ortega, 2000; Spada and Lightbown, 2008; Spada and Tomita, 2010). Given such positive results, it might be unsurprising that grammar instruction is often an expected part of foreign language curriculum.

Long (1991) compares *focusing on forms,* where lessons are structured to take learners through predetermined grammatical forms one at a time, and *focusing on form*, which originally meant that teachers reacted only to structures that learners were struggling with during communicative activities. These two common classifications seem to represent extremes of a continuum, however, which may be unhelpful. It is not unreasonable to think that there is a place for grammar instruction that is neither overly prescriptive (selected and taught in a predetermined order), as focus on forms seems to suggest; nor only reactive (as errors occur), as focus on form suggests. Research suggests that the form/ forms distinction may not be very important; Norris and Ortega's (2000) meta-analysis of instructional techniques found that there was little difference between *focusing on forms* or *focusing on form* in terms of instructional effectiveness, with both forms exhibiting a large effect size and showing a difference of only .07 standard deviations from each other. Perhaps as a result of this finding, little recent research seems to have been conducted on the issue.

Although Long himself argues that *focus on form* 'has always been, by definition, reactive' (Long, 2017:38), a review of research suggests that others have found this dichotomy to be too limited: focus on form has been broadened over time by some researchers to include pre-selected forms that teachers anticipated students would struggle with during the planned communicative activities (Doughty and Varela, 1998; Doughty and Williams, 1998; Ellis, 2001; Spada and Lightbown, 2008). Applying the broader view of focus on form, for example, teachers might conduct a needs analysis of their students' work to identify where problem areas are, thereby maintaining the reactive foundation of focus on form; and organise this information into lessons.

**Organising instruction of linguistic targets**

Johnson (1982:129) differentiated between *unificationist* and *separationist* styles of grammar instruction. He describes instruction and communicative activities in the unificationist style as being taught simultaneously, while in the separationist style, instruction focuses first on a grammatical structure, and then on communicative activities designed to promote use of the structure. A similar dichotomy is identified in the review of literature by Spada and Lightbown (2008), who use the

terms *integrated form-focused instruction (FFI)*, to describe grammatical instruction taking place within communicative activities, and *isolated FFI*, where communicative activities and grammatical instruction are kept separate.

Research suggests that both isolated and integrated FFI can be effective for L2 development. Spada, Jessop, Tomita, Suzuki, and Valeo (2014) investigated effects of isolated and integrated FFI on the passive construction with intermediate-level adult participants (mixed L1; aged 18-65) in a community ESL programme in Canada. There were a total of 109 participants from four intact classes, but only 67 of these completed all tests. Two of these classes received integrated FFI, and the other two received isolated FFI. The learners all received lessons delivered on the same topics and for the same number of hours (12 hours), and received both metalinguistic explanations and explicit CF. Participants were divided into two groups. Instruction on the passive construction for the Integrated FFI group (n=35) was given as an aside during communicative activities when learners made errors answering comprehension questions, while for the Isolated FFI group (n=32), instruction was separate from the communicative activities. For example, the Isolated FFI group completed a reading activity and practised with cloze worksheets using the targeted grammatical form.

Participants were administered two versions (version A and version B) of each instrument: a written error correction test, drawing on explicit knowledge; and an oral picture-description task, which, due to the time-pressured nature of the test, may draw on implicit knowledge (though the authors admit this is not a 'pure' test of implicit knowledge). The pre-test was version A of the instruments; post- and delayed post-tests were both version B of the instruments. Multilevel modeling was used to assess the results, finding that both groups showed similar statistically significant levels of improvement with both tests from pre-test to post-test and from pre- to delayed post-test ($p<.05$), but that the differences between groups were not statistically significant. Though without a control group, the researchers interpret the results to suggest that, as both isolated and integrated FFI groups statistically improved from pre-test to post-test, both forms of instruction can contribute to the development both implicit and explicit knowledge. These results suggest that, as Spada and Lightbown (2008) propose in their review of isolated and integrated FFI research, teachers may choose which form of FFI to use depending on their unique teaching context. They explain that isolated FFI may be preferred when, for example, the difference between the L1 and L2 language forms is great; when drawing learners' attention to features they may have failed to notice; or when learners are unlikely to encounter the forms outside of the classroom.

The taxonomies differ in details, but all 'share the common assumption that some kind of attention to form is necessary in L2 classrooms' (Nassaji and Fotos, 2007:15), a conclusion also shared by Norris and Ortega (2000) in their meta-analysis. That *attention to form is necessary* is also the conclusion drawn in the meta-analysis of 30 studies by Spada and Tomita (2010). The researchers categorised the targeted linguistic forms of the included studies as *Simple* or *Complex*, calculated by how many transformations were required to arrive at the targeted form; as this resulted in articles being classified as *Simple*, it might be argued, as the authors note, that different systems of classification might classify the targeted forms differently. It seems not to have been an issue in the current study, however, because regardless of the classification, 'explicit instruction was found to be superior in contributing to learners' explicit knowledge of [both] complex and simple forms' (p. 289f). The largest effect size was for explicit instruction of complex forms (d=0.88), and then for explicit instruction for simple forms (d=0.73). Implicit instruction produced much smaller effect sizes (Complex: *d*=0.39; Simple: *d*=0.33).

As suggested here, there seems to be a strong positive benefit for focusing learners' attention on targeted grammatical forms, and though there may be factors to consider when deciding whether to present the targeted forms in isolation or integrated into the lesson (Spada et al., 2014), both of these ways of teaching grammatical forms can contribute to learners' L2 development.

### 2.3.3 Section summary

Section 2.3.1 explained the theoretical underpinning of L2 development for this study. Firstly, skill acquisition theory suggests that if learners are given opportunity to practice declarative knowledge, this might lead to the proceduralisation of the skills being practiced. That learners need to work with the forms being practiced in a way that prompts deeper analysis of the grammatical points was discussed in light of ideas from human information processing theory. It was also discussed here that learners can become overwhelmed with information and not know where to focus their attention, thereby highlighting a role for the expert as one who can narrow the field of focus for the learner. Section 2.3.1.3 discussed the interaction hypothesis, which incorporates elements from the other two theories to create an understanding of the process of information being taken in and used by the learner, contributing to the understanding of how grammar instruction might promote L2 development. It was seen in this section that research demonstrates the value of form-focused instruction, and that the decision to provide isolated or integrated instruction may best left to the teacher, considering variables such as the complexity and salience of the form being learnt and the learners' proficiency (Spada and Lightbown, 2008; Spada et al., 2014).

Instruction alone might be an insufficient condition for enduring improvement of grammatical accuracy; skill acquisition theory suggests that practice is needed, but, as the next section will discuss, practice might be of questionable value if feedback is not provided (Anderson, 1983), leaving learners to practise inaccurate grammatical forms. Teachers can provide feedback to help learners gauge the accuracy of their output and improve it, possibly using explicit knowledge to do so. The next section will explain how corrective feedback (CF) contributes to L2 development and how the various types of CF are supported by the theories explained.

## 2.4 Corrective feedback for L2 writers

Cumming (1990; Section 2.2) found that L2 writers seemed to actively seek out errors and make their own corrections. Skill acquisition theory suggests this can be encouraged through the provision of feedback; Anderson (1983:407) sees feedback as essential to the acquisition of skill, as feedback can '[indicate] that a particular production has been misapplied, and… [can indicate] what the correct action should have been'. This section will provide a discussion about corrective feedback and factors which research suggests can influence its effectiveness, and is divided into three main sections. Section 2.4.1 will discuss various types of CF, how they are hypothesized to aid L2 development, and what research shows about each one. Next, Section 2.4.3 will examine factors which can influence the effectiveness of CF that are dependent upon the learner. Section 2.4.4 will discuss steps teacher might take which have been shown to influence the effectiveness of CF. Arguably, as *CF type* is learner-independent and is hypothesised to influence CF effectiveness the discussion of CF type could have taken place alongside the consideration of learner-independent factors; however, the centrality of CF to the discussion suggests it is best to discuss it separately from other factors.

The concept of error correction is likely well known to most researchers who study some aspect of L1 or L2 writing and to most practitioners. It has been argued that the term *error correction* 'seems to imply that feedback provided in response to learner error ultimately leads to the elimination of such errors' (Leeman, 2007:112). Given the prevalence of the term in the literature, however, it seems that most researchers accept it, however cautiously. Certainly, one may *hope* that error correction can help lead to the elimination of targeted errors, but the current study does not assume that error correction will necessarily result in complete elimination of targeted errors. To avoid misunderstandings due to terminology, a discussion of terms is next.

Chaudron (1977) writes that researchers of his time commonly defined corrective feedback as 'any reaction of the teacher which clearly transforms, disapprovingly refers to, or demands improvement of the learner's utterance' (ibid.:31). Almost thirty years later, the definition had remained largely unchanged: Oliver and Mackey (2003:519) described corrective feedback as a 'response to learners' non-target-like production'. Iwashita (2003:2) wrote that feedback was 'some kind of native speaker response to what the learner has said'. Making a slight modification of this, Afitska (2015:62) argues that 'native speaker' might more accurately be replaced with 'a more expert language user', but otherwise left the definition intact. Lightbown and Spada (2013:216) define CF as '[a]n indication to a learner that his or her use of the target language is incorrect'. Most recently written CF has been defined as 'a written response to a linguistic error that has been made in the writing of a text by an L2 learner.… [and the focus tends to be] on grammatical errors' (Bitchener and Storch, 2016:np). A close look reveals some differences between the definitions:

*Who is providing the correction?*

- the teacher (Chaudron)
- a native speaker (Iwashita)
- anyone with more language expertise than the learner (Afitska)
- not specified (Oliver and Mackey; Lightbown & Spada; Bitchener & Storch)

*Why is the correction being given?*

- to deal with a problem in the learner's production (Chaudron; Oliver & Mackey); specifically, a grammatical error (Bitchener & Storch)
- to indicate that the learner's 'use of the target language is incorrect' (Lightbown & Spada)
- not specified (Iwashita and Afitska). Note that as the term '*corrective* feedback' implies that there is a problem to be dealt with, it is possible that Iwashita and Afitska felt it was not necessary to add this detail. As Nassaji (2015:3) points out, '[a]ny kind of corrective feedback is in essence a response to a learner error…. [I]t is the error that triggers the feedback'.

*What is being corrected?*

- speech (Chaudron; Iwashita; Afitska)
- 'production'; that is, speech or writing (Oliver & Mackey)
- 'use of the target language' (Lightbown & Spada)
- writing (note: this definition was given for *written* CF; Bitchener & Storch)

It is Lightbown and Spada's (2013) definition of CF that will be used in this study because: this definition leaves open the issue of who is providing the correction; the CF focus can apply to both speaking and writing; and importantly, it holds that the purpose of CF is to indicate incorrect use of

the target language to the learner; CF should not merely offer a correction, but draw the learner's attention to the form so that noticing can occur, furthering L2 development. This is a central function of CF in the current study.

As Bitchener and Storch (2016) point out, grammatical errors tend to be the focus of written CF; and it is errors that prompt the CF (Lightbown and Spada, 2013). Errors have been described in the literature as 'idiosyncrasies in the interlanguage of the learner that are direct manifestations of a system within which a learner is operating at the time' (Brown, 2007:258); as 'those parts of conversation or composition that deviate from some selected norm of mature language performance' (Dulay, Burt and Krashen, 1982:138); and as 'the use of a linguistic item in a way, which, according to fluent users of the language, indicates faulty or incomplete learning' (Chun et al., 1982:538). Nassaji (2015:5) points out that 'a common feature of most definitions is that errors are language forms that do not accord with either the rules or the norms of the target language'; this can be seen in the definitions provided. As there is a common feature in the various definitions of *error*, as Nassaji (ibid.:5) explains, this common feature is applied to the current study: *language forms that do not accord with either the rules or the norms of the target language.*

Truscott sparked a lively debate when he used existing literature to conclude in his 1996 article that not only was corrective feedback *not* useful to L2 writers, but that it was actually harmful to them. He argued strongly that '[g]rammar correction has no place in writing classes and should be abandoned' (ibid.:361). While he, like many others, did not provide an exact definition of the construct, he seems to refer to a type of CF that matches the definitions above. Citing studies that had been carried out on CF, Truscott argued that CF was ineffective, and that studies which seemed to indicate its effectiveness were too flawed to be considered as providing evidence otherwise. Teachers were apparently shocked by this claim; Ferris (1999) writes that teachers 'recoiled as if they'd been punched in the stomach' (ibid.:2). She also revealed that she carried a secret hope that Truscott was right, and that teachers would have a reason to stop this time-consuming work. Researchers acknowledged the claims Truscott made in his original article regarding problems with the quality of feedback research; Ferris (2004) writes:

> I was struck by how different the various error correction studies were from one another.... Not only do they report dissimilar findings, but they are not even asking the same questions to begin with (ibid.:51f).

Researchers responded to the challenge of improving the design of feedback studies, and in 2004, Ferris wrote an updated view of the literature that had been carried out in the intervening years, which led her to conclude that corrective feedback is a 'necessary component of L2 writing

instruction' (ibid.:59). Gass and Mackey (2007:182) concur, explaining that feedback may 'help to make problematic aspects of learners' interlanguage salient and may give them additional opportunities to focus on their production or comprehension'.

Studies on CF tend to focus on two main themes: a) explicit vs implicit feedback, which concerns how discernible the correction is; and b) input-providing vs output-prompting correction, which concerns whether there is output required of the learner (Pawlak and Tomczyk, 2013). These distinctions form a taxonomy of written and oral CF Table 2-4 developed by Ellis (2009b:8); this taxonomy helps to clarify the different types of feedback and how they relate to one another.

**Table 2-4 Written and oral CF strategy taxonomy (reproduced from Ellis, 2009:8)**

| | *Implicit* | *Explicit* |
|---|---|---|
| *Input-providing* | Recast | Explicit correction |
| *Output-prompting* | Repetition<br>Clarification request | Metalinguistic explanation<br>Elicitation<br>Paralinguistic signal (gestures) |

Note that there is no overt differentiation between oral and written CF, though once a correction is written, it is obvious that a correction has been made; for this reason, Sheen and Ellis note that 'all written CF is necessarily explicit' (2011:594). Ellis concedes that it is an imperfect system because these CF tools are neither mutually exclusive nor necessarily distinct; for example, a teacher could offer an oral recast while also offering paralinguistic signals (e.g. gestures) as a cue. While it is clear how elicitation or paralinguistic signal could prompt output (e.g. by saying *There are two… **boy**?* or showing two fingers to prompt learner to use a plural form), it is less clear how metalinguistic explanation would necessarily result in output. Although Ellis (2009b) provides explanations for all other CF types, metalinguistic explanation seems to have been missed; an examination of work done by Ellis around the same time suggests that he might have been conceptualising metalinguistic explanations differently from common usage: he refers to metalinguistic explanations as *clues* (Ellis, 2009c; Ellis, 2010). While this conceptualisation of metalinguistic explanation might be less common, it does not seem problematic, as the learners still receive the explanation. Regardless, this taxonomy has been used in research as a way of understanding and discussing the theoretical assumption of each type of CF (Sheen and Ellis, 2011; Pawlak and Tomczyk, 2013; Shintani, Ellis and Suzuki, 2014).

### 2.4.1 Operationalising and measuring CF effectiveness

Before discussing the literature on CF types, it is useful to discuss how effectiveness is operationalised and measured. Researchers commonly operationalise CF effectiveness in two main

ways: (successful) uptake in speech or in revisions of written work; or a decrease in (targeted) errors in new writing. These two ways are discussed next.

Lyster and Ranta (1997:49) define uptake as 'a student's utterance that immediately follows the teacher's feedback and that constitutes a reaction in some way to the teacher's intention to draw attention to some aspect of the student's initial utterance'. They explain that uptake is useful because it 'reveals what the student attempts to do with the teacher's feedback' (ibid.:49), and they identify two kinds: '(a) uptake that results in "repair" of the error on which the feedback focused and (b) uptake that results in an utterance that still needs repair' (p49). With time, however, the definition seems to be transforming, with researchers now using uptake to refer to written corrections, and sometimes discussing *uptake* in terms of only *successful* repair (e.g. Storch and Wigglesworth, 2010; Santos, López-Serrano and Manchón, 2010; Hanaoka and Izumi, 2012).

Uptake suggests that a correction has been noticed, but it does not necessarily demonstrate presence or absence of learning. Nassaji (2011) found that adult ESL learners (n=42; various L1s; ages 18-35) in intensive ESL classes, who seemed to indicate their awareness that the teacher had made an oral correction by repeating the correction (a response which fits the description of (successful) uptake), later used the incorrect form again. Though this result is not measured with statistical tests, it is reasonable that uptake does not always result in learning. On the other hand, failure to exhibit evidence of uptake cannot be interpreted as definitive proof of a *lack* of L2 learning or acquisition (MacKey and Philp, 1998); however, sometimes the linguistic form learnt may not be what was intended. Yoshida's (2010) small study of Japanese L2 learners (n=7; English L1; some English-Chinese L1 bilingual) in a university in Australia audio-recorded learners interacting with their Japanese L2 teacher. Through stimulated recall interviews, it was revealed that although the teacher offered feedback, the learner had actually misunderstood which linguistic feature the teacher was commenting on. Thus, the learner felt that she had received positive reinforcement about a feature in her utterance, but it was not the feature targeted by the teacher. Uptake might not, on its own, be a reliable indication of participant understanding of CF.

The other way of operationalising CF effectiveness listed above is a decrease in errors; there are two ways of measuring errors. Researchers tend to use either t-units or error ratios to see whether there has been a decrease in (targeted) errors in unconstrained, free-production writing tasks (or *freewriting*; as opposed to *constrained free-production*, such as picture description tasks; Sheen, 2007). A t-unit, or *minimal terminable unit,* is defined as 'one main clause plus all the subordinate

clauses attached to or embedded within it' (ibid.:141). The concept of t-units was developed by Hunt (1965) as a method for more objectively analysing syntactical fluency in the writing of L1 English-speaking students in primary and secondary school, and is commonly applied to L2 research (e.g. Williams, 2004; Sachs and Polio, 2007; Evans et al., 2010). However, it is considered by some to be an inappropriate measure for some L2 writers. Gaies (1980) explains that it can be difficult for researchers to decipher the precise meaning of sentences written by lower-proficient students (a process called *re-punctuation* (Ney, 1966)), because the sentences are challenging to parse by t-unit or clause. If the writing of lower-proficient students is harder to decipher, it might necessitate a more subjective judgement on the part of the researcher than the writing of higher-proficient participants; this effectively differential treatment of writing samples might skew the results of the study.

A further problem with t-units, as Ney (1966) explains, is that they are unable to account for differences in sentence wording. For example, he points out that '*Then the rain falls and spring comes'* would be counted as two t-units while '*So he went on threw [through] the woods and pulled the feather out of his hat from the partridges* [sic] *and put a flower in his hat and walked [through] the woods'* would be counted as one t-unit (Ney, 1966:234). The increase in number of words in the second example clearly provides L2 writers with greater chances of more grammatical errors in that single t-unit, and this could influence the results.

The second common method of measurement is the use of error ratios. Noting the issues surrounding the use of t-units, as discussed above, some researchers have consequently chosen to use error ratios rather than t-units (e.g. Chandler, 2003; Truscott and Hsu, 2008; Van Beuningen, De Jong and Kuiken, 2012; Muñoz, 2014; Moradian, Miri and Hossein Nasab, 2017; Wang and Wang, 2015). Error ratios are calculated by taking the total number of errors and dividing the result by the number of words the participant has written, and then multiplying by, e.g. 100 to create an error ratio score per 100 words. Muñoz (2014:472) explained that this method offers 'the advantage of compensating for differences in [amount of] text or speech'.

This discussion of how CF effectiveness is measured in research will allow the reader to more fully appreciate the next sections, where a discussion of various types of CF and how they have been shown to be effective is provided.

### 2.4.2 Types of written CF

Having examined the benefits of writing, grammar instruction, and grammar correction, this section focuses particularly on what types of grammar correction are used and how effective research has shown them to be. A number of ways of delivering CF have been investigated, but not all are pedagogically feasible in all contexts. For example, there are *reformulations*, where a target-like version of a learner's writing work is provided to them (Thornbury, 1997). Reformulations could be problematic in settings where there are large numbers of students. The current study strives to maintain ecological validity, so despite any theoretical benefits, reformulations were not considered for the current study due to the time it would take to rewrite students' work. Another form of CF that was not appropriate was the provision of *models*, where learners receive a 'model' answer against which to compare their own writing (Hanaoka, 2007). Hanaoka (2007), described in Section 2.3.1.3, provided model answers for a picture description task, which provided some constraint on students' lexical and syntactic productions, facilitating the creation of a model answer. This might be useful in contexts where there are constraints on the writing task, but when participants are given freedom to develop their own ideas, creating models might be problematic. Thus, the discussion of written CF will focus on the following two common typologies which would have greater ecological validity in a variety of pedagogical contexts: a) *focused and unfocused* and b) *direct, indirect, and metalinguistic*.

#### *2.4.2.1 Focused and unfocused CF*

One way of classifying feedback is by how the feedback is being targeted. In *focused feedback,* tutors provide corrections only on certain, pre-specified error types, while in *unfocused feedback* they provide corrections on all or most errors. Relating this to the discussion in Section 2.3.1 about information processing, focused feedback could be hypothesised to be the more effective form for some learners, such as those with lower proficiency, as there are a limited number of items to focus on (Ellis et al., 2008). By this rationale, it would seem that unfocused feedback might overwhelm learners, but this does not always seem to be the case.

Evans, Hartshorn, and Strong-Krause's (2011) study examines a type of unfocused feedback called *dynamic feedback.* Hartshorn and colleagues had previously defined this as 'meaningful, timely, constant, and manageable' (Hartshorn et al., 2010:87). *Meaningful* means focusing only on learners' errors; *timely* means students receive feedback in the following class; *constant* means writing in every class; *manageable* refers to the amount of writing students are asked to produce.

Evans, Hartshorn, and Strong-Krause (2011) is a longitudinal study with a pre-test/post-test design, examining effectiveness of dynamic feedback with a small group of advanced-level university ESL students (n=30; variety of L1s; at least IELTS 6.5 proficient); it is largely a replication of an earlier study (Hartshorn et al., 2010), carried out to see if what had been effective in an intensive English programme (13 hours/week of class time) would also be effective in this non-intensive ESL writing course (number of hours not provided). Participants were divided into treatment and comparison groups; the treatment group received dynamic feedback, while the comparison group received feedback 'primarily on rhetorical aspects of writing, though feedback was also given on the linguistic accuracy' (p234). Both groups produced about 20 pages over the semester. The accuracy of the treatment group, measured by error-free t-units, significantly increased with a large effect size ($\eta_p^2$=.16); full statistical summaries are not provided. The scores of the control group, on the other hand, decreased, though not significantly. Note, however, that the information-processing theoretical discussion in Section 2.3.1 suggests that results might not be replicable with lower-proficient students, who might be overwhelmed by the volume of corrections.

It might be prudent to note that the large effect size reported in the article above could be the result of a misunderstanding. It is possible that the authors have attributed to their partial eta squared statistic criteria that are generally used for eta squared measures. Levine and Hullett (2002) note that while this conflation might not cause problems with some simple types of ANOVAs, this study applied a mixed model ANOVA to its data, which is a more complicated statistical test. Ferguson (2009) further points out that differences between the eta squared and partial eta squared statistics can be amplified when the sample size is small; there were only 15 participants in each group. Noting these words of caution, the effect size should be interpreted with care.

A problem with this design is practicality: busy teachers may hesitate to implement the correction regime of Hartshorn and colleagues. However, even if the issue of practicality can be overcome, positive results are not guaranteed, as demonstrated in Ellis, Sheen, Murakami, and Takashima (2008), who applied a more teacher-friendly definition. Comparing effects of correction of articles for first and second mention (e.g. *She saw a cat. The cat was cute.*) with effects of unfocused feedback, Ellis et al. (2008) examined these two feedback types in an EFL context with a 10-week longitudinal pre-test/post-test/delayed post-test design where the participants wrote short stories based on pictures. The participants were Japanese L1 university students (n=49), divided into three groups with eighteen learners in each treatment group (focused/ unfocused) and thirteen learners in the control group (receiving only general feedback on content). Increasing the level of practicality,

the definition of *focused/unfocused* differs in this study from more commonly applied definitions, as treatment groups received only four corrections, with the difference being that all four of these corrections were on articles for the focused group, while for the unfocused group, two were on articles and two were on grammatical structures other than articles. The authors estimated learners' proficiency to be intermediate based on previous learning history and university entrance exam results. As the tasks were controlled writing tasks rather than open ones, neither t-units nor error ratios were appropriate; instead authors calculated a ratio of obligatory occasions (where articles *must* be used) to actual correct uses. Results indicate that all three groups, including the control, interestingly, significantly improved grammatical accuracy of the targeted grammatical forms from pre-test to post-test, but ANOVAs reveal a non-significant difference between groups.

In addition to the unusual definition of focused and unfocused feedback applied in the study, the research design might have contributed to these results. Firstly, a ceiling effect in the pre-test suggests that the tests might have been too easy; there was, perhaps as a result, attrition of participants, leaving the final three groups with n=11, 13, 11. Secondly, control group participants were from a different class (a speaking class) than the two treatment groups (a reading class); it seems possible that, compared to the speaking class, the reading class might have studied grammar more intensively as they worked to understand the construction of the sentences they were reading. Sheen, Wright, and Moldawa (2009), without these problems, found that focused CF resulted in significantly greater improvement on targeted errors (articles) with their advanced-proficiency English L2 learners (postgraduate students in a US university; n=80). However, this seemingly positive finding might be questioned when considering that their practice-only group, which actually produced fewer writing samples than the treatment groups, also significantly outperformed the control group at the time of the delayed post-test. This indicates that other factors might have influenced the results.

It is difficult to draw conclusions about focused and unfocused feedback. A very limited focus on linguistic targets, such as only articles, as was seen here, is a common criticism of CF research; researchers argue that such limited focus cannot be expected to apply to more naturalistic settings where teachers provide instruction and CF on a much wider range of linguistic features (Xu, 2009; Crosthwaite, in press).

#### *2.4.2.2 Direct, Indirect, and Metalinguistic CF*

Another common feedback typology relates to the level of explicitness of the feedback being given: *direct* and *indirect feedback. Metalinguistic* feedback is also included in this category because it tends to occur alongside direct feedback, and is arguably even more explicit. Metalinguistic feedback, Bitchener and Storch (2016) explain, is an explanation of what caused the error (e.g. grammatical rules) in addition to the correction. For *direct feedback*, a knowledgeable person gives both the location of the learner-produced error and the correct form to be used, by marking the standard form on the paper; there is clear possibility for metalinguistic feedback to be given alongside direct correction. For *indirect feedback*, the knowledgeable person makes a mark on the paper signifying that there is an error, but does not offer the correct form. Learners receiving indirect feedback must identify the specific location of the error (if not indicated) and make the correction independently.

Ellis (2008:355) points out that the effectiveness of feedback is 'likely to depend on the current state of the learners' grammatical knowledge'. This would seem to indicate that CF might be useful for reinforcing learnt structures. Looking at these forms of CF from a human information processing perspective (Section 2.3.1.2), it could be hypothesised that indirect feedback could engage the learners in deeper processing, thereby helping them to better remember the information. However, even if teachers are able to provide indirect feedback, which, as the discussion will show, can be difficult to do, there might be barriers preventing learners from taking advantage of the opportunity to engage in this deeper analysis. For example, they might perceive the feedback negatively, feeling it is incomplete. They might be overwhelmed, and lack self-efficacy, feeling that they lack the knowledge necessary to handle the feedback. If so, then providing learners with the correction via direct feedback could make the task feel more manageable; the shortcoming of this is that it seems to fail to encourage the deeper analysis hypothesised by Craik and Lockhart (1972) as helping learners retain information. Metalinguistic CF might provide the right balance between direct and indirect CF. It is possible that when receiving direct CF, learners feel capable and thus have self-efficacy to handle the CF; and when the direct CF is accompanied by metalinguistic information, learners are prompted to engage more deeply. Metalinguistic information seems to appeal to the explicit knowledge of the learner, 'helping them to understand the nature of the error they have committed' (Ellis et al., 2008:356). The problem remains, however, as to how to help learners engage with the metalinguistic information they are provided with, rather than just glossing over it. These issues are examined next.

Studies indicate that teachers find it difficult to provide indirect feedback. In Ferris's (2006) study, teachers who had received instruction on providing indirect feedback and were asked to do so

actually gave direct feedback in 89% of cases; and a teacher-participant in Lee, Mak, and Burns (2016) began the academic year providing indirect WCF to her secondary school students, but over time, she abandoned the practice, citing time constraints and her perception of differing expectations of the institution, her peers, and her students.

Indirect feedback is, by definition, at least somewhat ambiguous, which can present problems for anyone receiving it – not just language learners. Undergraduate psychology students in Winstone, Nash, Rowntree, and Parker (2017) report that they dismissed tutors' indirect comments on their work because they felt unable to grasp what was intended. Faculty members report understanding and agreeing with students' frustration, explaining that merely telling students that there is a problem fails to help students acquire the skills needed to deal with the problem (Price et al., 2010). Understandably, then, language learners could also struggle when receiving indirect feedback, for they must not only see through the sometimes ambiguous comments, but in many cases must do so in the language being learnt. Students in Storch and Wigglesworth's (2010) study demonstrated difficulty in using indirect written CF. The participants were ESL postgraduate students with an advanced level of English proficiency (n=48). Participants completed a type of picture-description task appropriate to their advanced level (describing a graph). The researchers compared effects of teacher-provided reformulations to indirect feedback, which was apparently unfocused. The indirect feedback was provided using codes to indicate different types of errors. The researchers focused on uptake and retention across three writing sessions: an original writing (pairwork), a rewrite (pairwork), and new writing (individual). These results were not measured with statistical tests; instead, the main focus is on the case studies of pairwork. The researchers found that the students struggled to understand the indirect feedback, with participants debating amongst themselves during the pairwork about how to correct the error, and then sometimes both producing the same error individually in later writing.

Given these problems, it is unsurprising that other studies recommend caution when using this type of feedback with lower-proficient students. Illustrating the difficulty that students sometimes experience with deciphering indirect feedback is the following lower-level ESL student attempt at a revision for a sentence which had been given indirect feedback (underlining), as reported in Ruegg (2016):

> Original sentence: *After <u>graduate</u> from school, almost people work.*
> Feedback: You have several verb tense errors (underlined)
> Revision attempt: *After **graduated** from school, almost people **works**.*

Note that in addition to the underlining, the teacher provided an indication of what the problem was, thereby narrowing the field of possibilities of corrections for the student to consider. Despite this, the student not only failed to make the correction of the underlined word, they changed the other verb in the sentence, resulting in the number of non-standard grammatical structures in the sentence increasing from one to two (bold emphasis mine).

Li (2009) uses the results from his study on speech corrections (recasts) to advise that while more proficient students might respond well to both direct and indirect correction, lower-proficient students might better understand a more direct form of correction. Offering a possible explanation for this, Blau and Hall (2002) point out that when tutors attempt to indirectly elicit information from learners, the learners can become frustrated when they feel that the questioning is a guessing game. If students feel frustrated by such techniques even in more intimate one-to-one situations, it is understandable that having the indirect feedback marks on paper could produce even greater levels of frustration; indeed, Hedgecock and Lefkowitz (1996) reported student frustration trying to decode their written feedback and what the authors termed a *lexicogrammatical guessing game*. Difficulties understanding feedback is not limited to language learners. Glover and Brown's (2006) qualitative study of students at two different universities in the UK (n=147) and undergraduate students reported in Winstone et al. (2017) suggest that L1 learners can feel similar frustration about feedback received in their content courses. One interviewee explained that teachers' indirect feedback can '[k]eep you wondering, I don't wanna wonder. Just tell me what to do!' (p10).

Bitchener and Knoch's (2010) study further suggests that indirect WCF might be problematic for some students. Their participants were university students in an ESL environment (n=63) with various L1 backgrounds, determined by unspecified measures to be advanced writers of L2 English. Learners were divided into four groups: 1) indirect WCF, 2) direct WCF + written metalinguistic explanation, 3) direct WCF + written metalinguistic explanation + oral explanation, 4) control group, which did not receive any feedback. The grammatical target structure was articles for first and second mention; the task was a picture-description task. Like Ellis et al. (2008) (Section 2.4.2.1), who also used a controlled task instead of a freewriting task, errors were measured by ratios of obligatory and actual uses. All treatment groups (but not the control group) initially showed positive change, but with no statistically significant differences between groups; however, the indirect WCF group failed to maintain their improvement, and their 8-week delayed post-test results were statistically similar to those of the control group. In addition to the 2010 study, Bitchener and Knoch have conducted several other similar WCF studies using similar tasks, treatments, measures of analysis,

and study designs (2008; 2009a; 2009b), which produced similar results: treatment groups tend to improve, and demonstrate no significant differences between them from pre- to post-test. From these studies, it seems possible that direct feedback and indirect feedback (if learners are able to use it) can be effective in improving linguistic accuracy, though indirect CF might not have long-term benefits.

Using a similar design (i.e. task, measures, analysis) to the three studies discussed above, though, Sheen (2007) did find significant differences in her study, including between treatment groups. Like Bitchener and Knoch's studies above, this was in an ESL context (a community college) and the learners, again of various L1s (n=91) were divided into three groups: a) Direct CF; b) Direct CF + written metalinguistic explanation; and c) Control, who neither completed the tasks nor received feedback; the definitions of each of these conditions was similar to how Bitchener and Knoch had defined their similarly-labelled groups; note, though, that there was not an oral metalinguistic (grammar lesson) group. Using a pre-test – post-test – delayed post-test design, Sheen, like Bitchener and Knoch, focused on articles of first and second mention. Participants received two treatments of feedback on guided narrative writing tasks. Statistical tests suggest that both treatment groups improved significantly more than the control group, and the direct WCF + metalinguistic explanation group performed significantly better than the direct WCF group ($p<.05$; effect size not given), but only between the immediate post-test and the delayed post-test; for reasons that are unclear, the two treatment groups were not significantly different at the time of the immediate post-test. Perhaps the students were able to internalise the knowledge with the extra time given; another possibility is that the oral metalinguistic information took the form of a 15 minute full class discussion of the written meta-linguistic explanation that the writers wanted to have clarified' (p 212). Students (51%) and staff (42%) in Mulliner and Tucker (2017) report their perception that feedback to a group is not an effective way to deliver the feedback; perhaps learners in Bitchener and Knoch treated the metalinguistic information as a general lecture, and let their minds wander. Regardless, with treatments focusing on one linguistic form, it is difficult to draw conclusions about how this result might generalise to an L2 writing class where students are creating their own writing and feedback is given on multiple linguistic forms. It is relevant to note that these groups a) were in an ESL context, surrounded by the target language; b) comprised learners ages 21-56; and c) included a range of educational backgrounds, from high school diplomas to doctoral degrees. In addition, the groups in the study comprised intact classes. A discussion of group comparability is not provided, but with such a range of variables, it is possible that other learner factors influenced their performance.

A closer look at the above studies is warranted; the results in Bitchener and Knoch (2008, 2009a, 2009b) indicated similar results for the treatment groups; however, with a very similar design and in a similar context, Sheen's (2011) findings suggest significant differences between her treatment groups. It is possible to speculate on some reasons for this: Bitchener and Knoch used two-way repeated measures ANOVAs for two of the studies, and a mixed design ANOVA for the other. Sheen (2011), on the other hand, ran multiple one-way ANOVAs and followed these with two-way repeated measures ANOVAs. Multiple statistical tests can increase the amount of error introduced into the results, so it is possible that Sheen's (2011) results are influenced by statistical error. Other factors might have contributed to the inconsistencies between these similar studies: Sheen's participants had been in the English-speaking country where the study took place for one to five years while those in Bitchener and Knoch had only been in the country around 18 months. Another difference between the two studies is the estimated proficiency of the students; neither Bitchener and Knoch nor Sheen gave proficiency tests to their participants, instead gauging proficiency according to the class the participants were in. It is possible that the difference between Bitchener and Knoch's low-intermediate students and Sheen's intermediate students was just enough to create the differences seen.

Classroom teachers would likely find it unsurprising that studies such as the ones above have been criticised for using such limited study designs. Bruton (2009) argues persuasively for pedagogically sound methodological designs, such as studying multiple grammatical structures and assuring that the design would be possible for teachers to implement. Xu (2009) further criticises that the use of such limited grammatical structures creates transparency in the study, arguing that this could allow the participants to see the target of the study and focus their attention accordingly.

Using a more ecological design, Crosthwaite (in press) investigated the development of accuracy of tertiary learners (n=87; minimum IELTS 6.5; ages 18-19; mostly Cantonese L1) in one-term EAP courses in Hong Kong, concluding that 'neither direct nor indirect feedback on errors (nor online forms-focused instruction) is having the desired effect' (ibid.:22). There were a total of five sections of the course taught by two teachers – a Cantonese L1 teacher in her 30s and an English L1 teacher in his 50s. Assigned tasks contributed to participants' mark for their course. For Task 1 (which also functioned as the pre-test, taking place in Week 2), learners were given 90 minutes to write an 800-word report or essay (their choice; in academic genre). Following this, indirect feedback was provided on their work, presumably on all errors; from this, students selected eight targets on which

to focus during their independent studies (required Moodle modules). In Week 10, students submitted another 800-word report or essay, this time one they had prepared at home; CF was provided on this the following week. In the final week (14) of the term, students were given three hours to write a report/ essay in class; it is unclear what classroom activities took place between weeks 11 and 14, but regardless, because CF was provided in week 11, the final task cannot be considered a delayed post-test, and the author does not refer to it as such.

Teachers provided feedback as they wished throughout the term; it is unknown what kind of feedback and which errors were targeted – or whether any grammatical feedback was provided at all. All data were converted into a corpus, and the analyses were error ratios. It seems that the researcher conducted a Kruskal-Wallis test on the data, finding that there was a significant effect over time ($p<0.001$); to locate where the differences were, it seems that multiple t-tests were used. These post-hoc tests revealed significant differences between Task 1 and Task 3 ($t(2) = 47.554$, $p < 0.001$), and between Task 2 and Task 3 ($t(2) = 37.001$, $p = 0.004$). It is unclear why a Kruskal-Wallis test was selected instead of an ANOVA, and it is equally unclear why t-tests were selected for post-hoc tests. However, mixed-effects linear regression analysis suggests this difference to be the result of task type (essay/ report) or condition (in-class/ at home), ($F (1, 849.55) = 8.946$, $p = 0.003$). Interaction between task type and time did not predict error frequency significantly ($F (2, 654.71) = 1.11$, $p = 0.330$). While the design of the study is ecological, it also allows much room for confounding variables; a similar design with greater control would allow for more reliable conclusions to be drawn.

### 2.4.3 Learner-dependent factors influencing CF effectiveness

From the discussion of the different types of corrections, it is clear that studies are producing inconsistent results, with some studies reporting significance and others not. Nassaji (2016:540) writes that 'conflicting or contradictory results… indicate that the [issue] is complex and may vary depending on a number of factors'. This discussion will focus on some of those factors as they pertain to the study.

That no *one* CF technique is effective for all participants seems intuitive, and is also supported in the literature. Researching effects of error correction on L2 grammar knowledge and oral proficiency over the course of a school year with Dutch L1 high school students (n=35) studying French L2 in two classes, DeKeyser (1993:505) asked the teacher in one class to provide correction 'as frequently and explicitly as possible… [and] the other … to avoid correction as much as possible' (DeKeyser,

1993:505). Ten classes, selected by an unspecified method, were recorded, transcribed, and analysed, finding that in those ten classes, the students in the *Receiving corrections* condition received a total of 188 error corrections while the class whose teacher was avoiding corrections received merely 18 total error corrections. Examining pre-test and post-test results, the findings indicate that there was no main effect for correction, and that a number of variables may have contributed to this, i.e. previous achievement, grammatical sensitivity, motivation, and anxiety. Other factors might also contribute to the observed results: the small numbers of participants, the fact that they had two different teachers, and the fact that the two conditions were rather extreme – correcting as much and as frequently as possible, as the correction group teacher was instructed to do, might have overwhelmed the students with too much information, preventing them from learning from the corrections. Nonetheless, these results, combined with the above discussion, warrant a discussion of mediating factors which can contribute to differences in learning.

There is a wide variety of factors which have been identified as potentially influencing language development; the ones focused on here were selected upon the consideration of the aims of the study, the constraints of the institution, as well as practical matters, such as time. Mediating factors can be grouped into three broad categories: a) contextual factors, such as the learning environment; b) motivational/ affective factors, such as interest and beliefs; and c) cognitive factors, such as working memory and language aptitude (Bitchener and Storch, 2016: n.p.), though there could also be an argument for conceptualising *contextual factors* as part of *motivational/ affective factors* due to the influence the former might have on the latter (Kormos, 2012). The discussion in the sections below will draw from each category, keeping in mind the research questions and the population, creating a balanced consideration of the mediating factors impacting on learning. Perhaps, in part, because much research is conducted in mixed-L1 environments, contextual factors seem less focused on in research. Section 2.4.3.1 will present research on contextual mediating factors in order to situate the discussion of the context of the participants in the current study. The next section is called *Learner preferences*, and includes interest, which draws from the motivational/ affective category, but as learners' interests might lead them to certain kinds of learning environments, it has a contextual element to it as well. Self-efficacy, from the motivational/ affective category, and language aptitude, from the cognitive category, are often considered to be major contributing factors which can influence language learning, as will be discussed in Sections 2.4.3.3 and 2.4.3.4, respectively.

#### *2.4.3.1 Learners' context (prior and current) and experiences*

Previous educational contexts can establish expectations for learning, which might affect how a learner engages with a given pedagogy; an example could be students moving from secondary school, where there might be more direct guidance, to university, where they might be expected to be more autonomous. Orsmond and Merry (2011) analysed essay feedback on biology assignments for undergraduates (n=19) and then interviewed students to find out what they thought of the feedback and how they understood it; this led them to say that 'there was an implicit assumption on behalf of the tutors that students would know how to apply their feedback to future work' (ibid.:131). They speculate that one reason for the apparent inability to apply feedback may be because the learners are inexperienced in interpreting feedback independently, perhaps having received more guidance in their secondary schooling: all but five students interviewed reported having such problems. In discussing their feedback with the researchers, students overvalued trivial corrections while skipping over comments which detailed larger problems. One student felt that tutors always criticised his spelling, but when researchers reviewed this student's writing, no spelling corrections were found; there was, however, an important comment regarding a missing part of the essay which the student failed to notice. The authors suggest that inexperienced learners often have limited understanding of tutor comments, and that tutors need to work with students to assist them in learning how to use feedback for future improvement.

Of particular pertinence to language learners, one way of describing previous educational experiences is by the language learning context: whether the target language (English, in this case) is a majority language spoken in the area where it is being learnt, as in the U.K. (ESL) or not, as in China (EFL). It is acknowledged that not all English-language learning contexts fit neatly into this dichotomy, as Nayar (1997) rightly argues; and even within ESL contexts, there is a wide variety of learners: those who are there only for a brief period; those who moved there earlier (or later), etc. (Ferris, 2009:6ff). Heeding this advice, care will be taken in this discussion not to overgeneralise.

Language learners represent a vast number of different cultures. Spencer-Oatey (2000:4) provides a definition of the complex concept of culture:

> Culture is a fuzzy set of attitudes, beliefs, behavioural conventions, and basic assumptions and values that are shared by a group of people, and that influence each member's behaviour and each member's interpretations of the "meaning" of other people's behaviour.

Different culture groups adopt different educational systems, and it might be unsurprising that research suggests that a given pedagogy might not be equally successful in all different educational systems; logically, CF techniques might be similarly affected by context. Examining possible reasons

why a given pedagogy in one region might be less successful in another, O'Sullivan (2004) suggests that culture plays a part. 'Learner-centred approaches were developed in the West and are appropriate to the Western focus on the individual' (O'Sullivan, 2004:595).

Williams (2017:28f) writes that 'most honest observers will note that Chinese, Japanese, and Korean cultures have vastly more similarity to each other than to, say, African cultures or European ones'. Teachers from Western countries teaching in East Asia can be heard commenting '"[t]he students are so quiet! They never want to speak during class. How can I get them to participate more in class activities?"' (Williams, 2017:47). Faculty in Western institutions can be heard making similar comments about the East Asian students in Western classrooms (Marlina, 2009). A possible – and common – explanation for what has become known as *the passive Asian learner* (Cheng, 2000) is that these students come from Confucian Heritage Cultures (CHC). Attributes commonly associated with CHCs in the literature include values such as group harmony and face; the teacher is the main and final authority; that uncertainty is to be avoided (Williams, 2017). If this is the case, it might be understandable that some Western pedagogies might be less effective in this context than others. However, when Wong (2004) investigated 'the common view that [Asian international students'] cultural background is some kind of stumbling block for quality learning in a western higher education system', he found that although students required a period for adjusting to the new learning expectations of the Australian university, this could be accomplished in a few months. This suggests that it might not be culture *per se* that is the problem, but rather, that it is students' prior learning experiences and lack of familiarity with the pedagogies that is causing problems (Cheng, 2000:141f). If so, as Wong's study suggests, giving students some time to adjust to the new pedagogies could solve the problem. Other possible reasons can be found in the literature as well, including language proficiency, content knowledge, personality, and anxiety (Lee, 2009; Marlina, 2009; Tran, 2013).

This is not to suggest that culture does not play a role at all; unfamiliar cultural habits and expectations may compound other problems. However, as suggested in Wong (2004) above, this can be overcome. Postgraduate ESL students (n=13) in Tian and Lowe's (2013) study were so upset after receiving CF that they refused to read the comments. The learners expressed feeling *frustrated, depressed, demoralised*, and *heartbroken*. Advanced ESL academic writing students in Best, Jones-Katz, Smolarek, Stolzenburg, and Williamson (2015) echoed the same sentiments, expressing anxiety. One added in an interview that they even avoided looking at feedback because they were 'afraid it will destroy my whole day' (ibid.:11). One reason for such negativity regarding feedback

might, according to Higgins, Hartley, and Skelton (2001), be that the tutor plays a dual role for students, as both helper and judge, so students may feel betrayed. Fortunately, by the end of the year, the students in Tian and Lowe's (2013) study reported accepting the feedback and appreciating its value. Perhaps continuing to receive similar CF from multiple sources, and gaining experience in the new academic culture allowed them to see that tutors were not being overly harsh. These participants were learning in a target-language-rich environment; the limited contact and short time that students in EFL contexts might have to make this adjustment to unfamiliar pedagogies may be insufficient, leaving the students with negative feelings.

Interestingly, the strong negative affective reaction to feedback seen above is not corroborated by all studies. For example, in Elwood and Bode's (2014) research with a large multi-cohort study of first-year EFL students (n=410) in a required writing class in Japan, the researchers found that the participants, who came from a variety of majors, reported fewer strong negative affective reactions to feedback than positive ones. For example, on a 5-point Likert scale (1 = Strongly disagree; 5 = Strongly agree), feelings of depression and shock were rated 2.23 and 2.4, respectively; on the other hand, positive affective reactions scored higher on the scale: feeling happy (3.73) and motivated (4.12). It is possible to reflect on why these students did not exhibit the negative emotions associated with feedback that have been seen in other studies, despite the fact that their course was also a required one. One possible explanation for the difference might be L2 proficiency and/ or confidence. Only the highest-scoring students were admitted into the writing class where the study takes place; perhaps this elite experience helped students feel confident regardless of their feedback. In addition, most of writing done for the class was rather informal (a memo, an email, an internet chat, a self-introduction), with students writing about their thoughts and experiences. Perhaps this familiar territory allowed the students to relax and remain confident in the face of CF. If all of the writing had been academic in style, as in Tian and Lowe (2013) and in Best, et al. (2015), the students in Elwood and Bode's study might have felt less comfortable; researchers report that the students rated writing a 5-page report and a speech as high-anxiety events (3.83 and 4.17, respectively, on the same 5-point Likert scale described above), compared to how they rated informal styles, such as writing a memo (2.13) or a self-introduction (2.70). It is feasible that had these students been writing academic-style assignments, the unfamiliarity and the anxiety might have caused them to report feelings in line with those in Tian & Lowe (2013) and Best, et al. (2015).

That students have negative reactions to feedback should not be interpreted to mean that these students did not wish to receive feedback. Given the number of studies reporting that students do

value feedback (I. Lee, 2008), it seems more likely that the students would have simply felt happier had their feedback been positive (understandably). That students value feedback has been well documented (I. Lee, 2008), and can be summarised by the following quotation from a student in a university ESL writing class: 'If the professor doesn't make any changes or tell me what's wrong with it, what's the point of coming to [class]?' (McMartin-Miller, 2014:31).

#### *2.4.3.2 Learner preferences*

University English courses can contain students from a variety of majors. The various interests and aptitudes that lead students to these majors may affect how students perceive and process feedback. Wong and Nunan (2011), for example, found that their participants, classified as less effective L2 learners, tended to indicate preference for passive learning techniques. The researchers used a 30-item Likert questionnaire to examine self-reported learning techniques of two groups of undergraduate EFL students (apparently Chinese L1), divided according to their self-reported English grades on a university entrance exam. Those who self-reported having received an A were categorised as *more effective* (n=77) and those who self-reported an E or F (n=33) were categorised as *less effective*, and those reporting mid-range grades were not included in the study so that a clear division could be made. On the survey, participants classified as *less effective* tended to identify English learning techniques that required more teacher input. Interestingly, these learners came mostly from the field of science. Though it may be problematic to use self-reported exam scores as measures of learning effectiveness, the study nonetheless indicates the possibility that the interests and preferences that lead students to a particular field of study might also result in differences in how students learn.

Why was it mostly science majors who were 'less effective' English learners in this study? It is possible that these learners, having a science focus, had lower interest in English; if interest is low, the learners may be engaging in Craik and Lockhart's (1972) Type-I processing (e.g. memorising), rather than the deeper (Type-II) engagement that can result in better learning. Possible reasons for differences in science majors can be found in the literature. For example, the students from disciplines where linguistic accuracy is less emphasised might be less interested in the linguistic accuracy of their own work. A Chinese L1 participant from the field of Engineering in Hong Kong in Hyland (2013b:182) explained that in their field 'professors would not give us feedback on [incorrect tense usage]. They would give feedback like the units, abbreviations, calculations, etc.' Another participant pointed out that writing for their field (Computer science) required only technical language: 'what I have written is related to the technical knowledge and so I can easily do it. I can finish the task by copy and paste. I can write up the assignment without thinking'.

Research also suggests that differences in learning styles might contribute to the differences between learners in Wong and Nunan (2011). For example, Rao and Liu (2011), asked Chinese L1 EFL university students (n=40) to track their English study methods in diaries, and, focusing only on the most and least mentioned activities (≥14 or ≤7 times), examined how students studied English, and classified respondents into natural and social science majors. The results, though not statistically tested, suggest that social science majors mentioned participating in more reading, practising, and revising types of activities; moreover, they reported enjoying these activities. Science majors, though, mentioned analysing and memorising more often, and noted that the activities were more necessary than enjoyable. Human information processing theory would hypothesise *memorising* (Type-1 processing) to be a less effective learning technique.

With only forty participants, it could be argued that Rao and Liu (2011) lacked sufficient participants to allow for any conclusions to be drawn, but larger studies seem to corroborate these findings. Gu (2002) found, with a larger sample size, similar results regarding vocabulary learning: science majors (n=347) and humanities/social science majors (n=298; females comprised approximately half of both groups). Using t-tests, Gu (2002) compared questionnaire results from the two groups. Though there were many similarities, significantly more science majors reported memorising ($t=-2.03$, $p<.05$) and analysing ($t=-2.33$, $p<.05$) than did the social science majors (*df* unreported). Unfortunately, the use of multiple t-tests might have introduced error into the results, but combined with Rao and Liu's (2011) study above, it seems possible that the methods science majors use to study English may make them more receptive to certain kinds of WCF than other kinds. For example, if science majors tend to find analysing sentences useful for language learning, perhaps they will respond well to feedback which makes use of analysing grammar. Interviews with undergraduate participants (n=28) in McLean, Bond, & Nicholson (2015) offer another possible reason for differences: how students themselves conceptualise CF; for example, some might perceive it as providing the right answer, as a general guide, or as offering alternative perspectives. If mathematics students, for example, are accustomed to receiving corrections of 'right' answers, it is possible that this expectation could carry over, influencing how the students expects to receive CF in their L2 writing classes.

The pattern seen in the above studies indicates that it might be worthwhile to consider the possibility of differences in learning preferences between majors and to invest in further study, though other possible explanations for differences exist. For example, neither study tested learner proficiency; if the science majors were simply lower in proficiency than the social science groups,

this might have contributed to the observed differences. Indeed, Esteban and Roca de Larios (2010) suggest that lower second language proficiency may explain difficulties learners experience with grammar correction. Their participants were Spanish L1 adolescents with lower-intermediate English L2 proficiency (n=17; aged 15; apparently EFL). When given a model answer and asked to use it to edit what they themselves had written, the students focused almost exclusively (83% of corrections made) on lexical differences between the two texts, and largely ignored problems of syntax. This is unsurprising considering the findings of Gass et al. (2003), as discussed in Section 1.2, who also found that without explicit instruction, problems of syntax went unnoticed by learners. Esteban and Roca de Larios (2010) suggest that this could have been due to low proficiency; indeed, if the models made use of grammar structures beyond the ability of the learners, they might have ignored the grammar and focused on what they could more easily understand. Thus, direct WCF might have produced different results. Note, though, that the learners were aged 15; learning to use feedback may be a matter of maturity. Williams and Kane's (2009) analysis of data from existing satisfaction surveys from twelve universities suggests that younger adults (aged ≤24) may not understand the value of feedback; older survey completers (aged ≥25) reported finding feedback to be more useful and more valuable than younger completers. While the difference between 24- and 25-year-olds is small, the data are from multiple years of surveys from twelve universities, offering some credibility. Though results were not tested statistically, it seems reasonable that learning to use feedback could be a skill that develops with age or experience. Less experienced students may require help from experts who can facilitate the process of their cognitive engagement with feedback.

#### *2.4.3.3 Learner self-efficacy*

Learners' perception of their own abilities constitutes another individual difference. Given two learners, one learner might perceive their own ability to be greater or less than that of the other, even if they actually possess the same level of ability. Belief about one's ability to successfully complete an activity is what Bandura (1977) refers to as *self-efficacy*. To Bandura, this differs from *confidence*, as confidence is not targeted at a specific domain (Bandura, 1997:382). Other researchers agree that self-efficacy and confidence are separate constructs, with the difference argued as being one of timing (Cramer, Neal and Brodsky, 2009; Morony et al., 2013). '[T]he self-efficacy question is asked prior to a (hypothetical) cognitive act and confidence is asked immediately after answering a cognitive item (Morony et al., 2013:81). Bandura (1977) points out that if one believes one can succeed in a given task, one may work harder to achieve it; this has clear relevance to academic settings. For example, lower self-efficacy could influence academic behaviours negatively, prompting what Lee, Bong, and Kim (2014) call *maladaptive achievement strategies* (e.g.

cheating on academic work), and higher self-efficacy could have a positive effect on learning. Hsieh and Kang (2010) examined Korean secondary school students' (n=192, 52% female) self-efficacy for attaining a given score 'on their next test'; this presumably refers to a test in their English L2 class. The authors report a Cronbach's alpha coefficient of .85 for internal consistency, but the instrument is not described. A comparison of the few details they provide with the scale items in the literature they cite for their own scale (i.e. Wood and Locke, 1987; Stajkovic and Sommer, 2000) suggests that the items in the instrument likely resemble those in Figure 2-1.

**Figure 2-1 Example self-efficacy instrument items**

| Item | Yes/ No | Level of certainty (0-100) |
|---|---|---|
| 1. I can attain a 60% on my next test | | |
| 2. I can attain a 70% on my next test | | |

Hsieh and Kang interpret their results as indicating that self-efficacy is a strong predictor of academic achievement in English L2 ($R^2$=.24, $F(1, 186)=58.67$, $p<.001$), an effect further strengthened by including *attribution*, or how learners credit their academic successes and failures. However, it is unclear whether *attaining (a given score) on my next test* is sufficiently specific for measuring self-efficacy. The self-efficacy measure appears to have been administered several months into the school year, so if the teacher tends to administer tests which follow a predictable template, then perhaps the instrument is unproblematic; however, it is unclear whether this is the case, meaning that participants could have a difficult time estimating their self-efficacy. The self-efficacy items in studies cited in support of this instrument seem to be more concrete, as depicted in Figure 2-2. The implications of this will be discussed later in this section.

**Figure 2-2 Two extracts of self-efficacy instrument items from the literature**

'The proportion of facts and concepts covered in [this] course that you feel you are able to memorize and recall on demand' (Wood and Locke, 1987)

| (Example) Items | Yes/No | % Confidence |
|---|---|---|
| 1. 60% of facts and concepts | | |
| 2. 70% of facts and concepts | | |

'Number of uses in 1 minute' (Stajkovic and Sommer, 2000)

| (Example) Items | Yes/No | % Certainty |
|---|---|---|
| 1. I can list 2 uses in 1 minute | | |
| 2. I can list 4 uses in 1 minutes | | |

The studies above look at self-efficacy as it relates to academic work more generally, but it is worth considering how self-efficacy might relate specifically to writing. Prat-Sala and Redford (2012) used a 7-point Likert-style self-efficacy scale for reading and writing (developed for an earlier (2010) study), and administered it to first- and second year undergraduate psychology students at a UK university (n=145; mostly female; ages 18-43). Though neither study reports reliability or validity measures for this instrument, Pearson's correlation tests performed on mean self-efficacy scores indicate that writing self-efficacy significantly predicted writing performance for both first year (r=.382; n=91; p<.001) and second year (r=.426; n=54; p=.001) students (effect size not reported). It is unclear exactly when the students completed the self-efficacy test, though the authors state that it was after the first-year students had completed their second assignment. Note that the learners were divided between four course tutors, and the tutors marked student work independently, so it is likely that the scores being given to the students reflected somewhat different expectations despite there being a shared general rubric dictating that the marks be based on a) content, b) criticality/ evidence, and c) general organisation/ mechanics. Not only might the marking have been different, but it seems likely that four different tutors could have four different ways of providing feedback, which could influence self-efficacy.

Ekholm, Zumbrunn, & Conklin (2015) examined the relationship between feedback, writing self-efficacy, and writing self-regulation with undergraduate students (n=115; mostly female) at a US university. Using three instruments, a) the *Writing Feedback Perceptions Scale;* b) an adapted version of a *Writing Self-efficacy Scale;* and c) a researcher-designed *Self-regulation Aptitude Scale*, (that is, students' beliefs about their ability to 'strategically manage the writing process' (p. 198)) researchers applied linear regression, finding significant relationships between self-efficacy and self-regulation ($F(1, 97) = 13.11$, $\rho < 0.001$; $R^2 = .11$; $\beta = 0.35$), where about 11% of variation was accounted for by self-efficacy; note, though, that measures of reliability and validity are not provided, neither in the current study, nor in the one upon which they base the development of the current self-efficacy instrument (i.e. Bruning et al., 2013). Assuming the reliability and validity of the instrument, it is noteworthy that Ekholm, et al. also found that students' self-efficacy had a significant influence on student perceptions of written feedback ($F (1, 110) = 10.09$, $p < 0.01$; $R^2 = 0.08$; $\beta = 0.29$), with about 8% of variation being explained by self-efficacy. When researchers controlled for self-efficacy, student perceptions of feedback demonstrated a significant positive impact on self-regulation ($F (2, 96) = 15.24$, $p < 0.001$; $R^2 = 0.23$; $\beta = 0.38$). Given these results, the authors encourage teachers to help students develop a positive view of feedback. Given the significant correlations between the three constructs, however, it seems that it would also be

important to help students learn to self-regulate their writing and to provide classroom experiences which show students their progress, which might lead to increased self-efficacy, such as doing revisions or helping students note their progress.

As this section might suggest, self-efficacy is a complicated construct, often misused or mislabelled in the literature; Bartimote-Aufflick, Bridgemean, Walker, Sharma, and Smith (2016), in their review of 64 studies on self-efficacy, found that the construct is often 'confused or conflated with other constructs seen as similar, such as self-evaluation, confidence, self-concept, and self-esteem' (ibid.:14). The discussion in the current section indicates that this might have been a problem in the 1980s in Wood and Locke's study (Figure 2-2), but the problem of conflating constructs can be seen in more recent literature as well (e.g. Lee, Bong and Kim, 2014). Adding to the challenge of using *self-efficacy* in research, sometimes measures being called *self-efficacy* in recent literature have borrowed self-efficacy concepts and instruments from research which may not have been applying the same understanding of the construct as it is now understood. For example, Joo, Lim, and Kim (2013) investigated whether achievement could be predicted by, among other things, self-efficacy (details not provided; presumably *general academic self-efficacy*) and task value, and interpreted their results as indicating that both self-efficacy, $\beta = .15$ ($t = 4.344$, $p < .001$), and task value, $\beta = .20$ ($t = 5.524$, $p < .001$), significantly influenced achievement. Joo et al. applied Bandura's understanding of the construct to their research, but the measure they administered to their participants was from Pintrich and De Groot's (1990) Motivated Strategies for Learning Questionnaire (MSLQ). The authors of the MSLQ state that the items for the MSLQ were adapted from even earlier studies; moreover, they do not mention Bandura or attempt to define self-efficacy, indicating a possible mismatch in definition and measurement. If this is the case, there are clear limitations on what interpretations can be made from the results.

#### *2.4.3.4 Learner language aptitude*

Another individual difference is learners' aptitude for language learning. Language aptitude has been described as the 'ability to reflect on language and extract abstract information' (Ellis, 2004:251), an ability which is described as relatively stable (Skehan, 1998:187f). DeKeyser (1993) calls language aptitude one of 'the most consistent predictors of achievement' (p. 504), and indeed, reviews of work in the field suggest that language aptitude can correlate with second language acquisition (Spada, 1997; 2011; Li, 2015). It is useful to identify which aspects of language aptitude are being considered. Carroll (1990) identified four components of language aptitude: phonetic coding ability; grammatical sensitivity; memory abilities; and inductive language learning ability. However, arguing

that grammatical sensitivity and inductive language learning ability are interrelated constructs, Skehan (1998:201) combines these into one construct called *language analytic ability* (LAA), which he defines as 'the capacity to infer rules of language and make linguistic generalizations or extrapolations' (ibid.:204). This restructured understanding of language aptitude has been accepted in research (Erlam, 2005; Sheen, 2007; Spada, 2011), and S. Li (2013) indicates his view that grammatical sensitivity and language analytic ability are synonymous terms. Theoretically, this links with the discussion of explicit knowledge as analysed knowledge in Section 2.3.2.1. As *analysed knowledge* is an awareness of grammatical rules, having an aptitude for seeing rules and patterns in grammar could influence one's ability to put use this knowledge advantageously when learning a language; Ellis (2004:251) sees potential for this to be 'the essential ability underlying the development of explicit knowledge'.

Though the relevance of LAA to language learning generally has been frequently investigated in the literature, few studies have examined its effect on the effectiveness of corrective feedback. DeKeyser (1993) investigated the effects of, among other things, LAA on CF effectiveness, as measured by written grammar tests and oral fluency and accuracy pre- and post-tests, with Dutch L1/ French L2 final-year secondary school students (n=35) in two classes taught by two different teachers. LAA was measured with the *Words in Sentences* section of the *Modern Language Aptitude Test* (MLAT; Dutch L1 version), a section requiring students to solve various syntax-related problems. One of the two classes served as the treatment group, and the other as the control. For an entire academic year, the treatment group teacher provided explicit, unfocused oral correction while the control group teacher made no special effort to give corrections. ANCOVA tests reflected a lack of significant interaction between treatment and grammatical sensitivity for any of the three measures (written grammatical test; oral fluency and accuracy tests; note that despite the classroom setting for this study, these tests were not classroom-based tests). Recall, though, that there were only 35 participants in the study altogether, leaving too few in each group to allow for statistical claims in either direction to be made.

Sheen (2007) examined the correlation between LAA and effectiveness of WCF with Korean ESL university students (n=91) in the US. Details of this study were reported in Section 2.4.2.2, so the discussion here will focus on her findings as they relate to LAA. A repeated-measures ANCOVA, with analytic ability as the covariate, suggested that students' LAA, as measured by a specially-designed test (Schmitt, Dörnyei, Adolphs, & Durrow 2004), affected their test scores over time ($F$ (2, 88) = 4.95,

$p < 0.05$). However, the same cautions apply to these results as to the ones reported on the efficacy of the feedback. There is much potential for confounding variables to have influenced these results.

S. Li's (2013) learners were Chinese L2 students (n=78; 2nd-4th years of language study; mostly English L1) at US universities divided into three groups – explicit CF (metalinguistic feedback), implicit CF (recasts), and control (no correction). An English version of the same section of the MLAT as in DeKeyser (1993) was used as a measure of grammatical sensitivity. Classifiers were selected as the target for correction; implications of this decision are discussed below. In three one-to-one sessions with the researcher (under experimental conditions), participants took written grammaticality judgement (GJ) and audio/oral elicited imitation (EI) pre-, post-, and delayed post-tests. For GJ, participants were prompted to mark written sentences as grammatical or not, and if not, to offer a correction; for EI, participants heard a statement about a possible everyday life experience (e.g. *I bought a jacket yesterday*), and were prompted to say whether the statement was true for them and then to repeat the statement. In the second of the three sessions, feedback according to group was given on the elicited imitation task, followed by the immediate post-tests; one week later the delayed post-tests were administered. Regression with gain scores and group indicates that both treatment groups significantly outperformed the control group, but a significant relationship with grammatical sensitivity was found only for the implicit CF group on the GJ test gain scores, $\beta = .44$, $p<.05$ (explaining 20% of variance). Before drawing any conclusions, though, a few details should be considered. Though it would be possible to have an academic argument about whether Chinese classifiers are syntactical or lexical items ((X. Li, 2013); a different Li), it is unlikely that this is a linguistic level that these particular language learners had achieved in their two to four years of instructed foreign language instruction. It seems more likely that they were working with classifiers as lexical items; for example, learning (and trying to remember) that the classifier *ke* is used with trees, and the classifier *tiao* is used with rivers (ibid.:17). If so, this has two implications: firstly, when operationalising language analytic ability for the study, Carroll's (1990) *memory ability* might have been more appropriate. Secondly, it was seen in Section 2.3.1.3 that learners independently tend to notice lexical corrections more than grammatical ones, so the results of S. Li (2013) might not be applicable to studies focusing on CF for improving grammatical accuracy.

Shintani and Ellis (2015) investigated CF and LAA with Japanese L1 university learners (n=118; 18-21) from a variety of majors in English L2 classes divided into four groups: Direct CF (with and without revision), and Metalinguistic explanation (with and without revision); there was no comparison or control group. The task was a dictogloss; the two targeted linguistic forms were a) indefinite articles;

and b) past hypothetical conditional. Language aptitude was measured with a Japanese translation of Pimsleur's (1966) Language Aptitude Battery. Learners were given one instance of feedback. Interestingly, correlations with overall accuracy score are not provided – only the breakdown of scores by targeted linguistic feature is given. There was no statistically significant correlation between aptitude and the delayed post-test scores for accuracy for any of the groups, and the immediate post-test showed mixed results:

Significant correlations for articles and LAA:

- direct feedback group who did NOT write revisions (r = .347; p=.030)
- metalinguistic group who did write revisions (r = .477; p=.006)

Significant correlations for conditional and LAA:

- direct feedback group who did write revisions (r = .482; p=.013)
- metalinguistic group who did NOT write revisions (r = .311; p=.033)

Shintani and Ellis suggest that those revising may simply be copying work without critically analysing it. They also suggest that different linguistic features may make different demands on learners' LAA. As it is difficult to draw clear conclusions from existing studies about the role of LAA on the efficacy of CF, there is a clear need for further work.

### 2.4.4 Teacher-controlled factors influencing CF effectiveness

It might seem incomplete to discuss personal factors influencing learners, as was done in Section 2.4.3, without also discussing here personal factors which can affect tutors. Surely, tutors are susceptible to being influenced by their prior learning and teaching experiences. While research does indicate that such experiences can influence beliefs and practices (Junqueira and Kim, 2013; Moodie, 2016), there is also evidence that quality second language teacher education programmes and reflective teaching practices can equip teachers with awareness and knowledge to counter the effects of their personal experiences (Vélez-Rendón, 2002). Thus, the discussion will continue with the assumption that the teacher is a qualified, reflective practitioner, and as such, the focus will be on how teachers provide CF: the clarity of the CF, whether they confirm that the CF has been understood, and the facilitation they offer students in using the CF.

#### *2.4.4.1 Clarity of CF*

Regardless of the CF type selected by tutors, if students are unable to read it or make sense of it, the feedback could be less useful, thereby decreasing its effectiveness. Hattie and Timperley (2007) persuasively argue that feedback must function to close the gap between students' *current* level of

understanding and the *desired* level of understanding. If written comments are unclear to students, they cannot contribute to the closing of this gap; according to this description, such comments cannot function as *feedback*. Sadler (1989) refers to feedback that is not useful to the learner as *dangling data*; the information is there, but it cannot be used to close the loop and aid understanding. The issue of dangling data can be caused by a number of different things. For instance, the problem could be one as basic as handwriting. Thirty percent of undergraduate respondents in L1 content courses in Robinson, Pope, and Holyoak (2011) reported that their feedback was sometimes illegible. Clearly, the problem could be compounded if the learners were trying to decipher poor handwriting that was also in the learners' L2. Thirteen of Ferris' (1995) 155 participants reported struggling to read tutors' handwriting, and an unspecified number of L2 learners in Carless (2006) also reported difficulty deciphering tutor handwriting.

Another possible cause of dangling data is when students are unable to understand what is meant by the tutor. Both L1 content and L2 learners report struggling when the meaning of the feedback was unclear. Participants in Ferris' (1995) study specified dissatisfaction with teachers' use of technical terms to describe their grammatical problems (e.g. *fragment, verb tense*) and with methods used in the provision of indirect feedback, such as circling; an interviewee in Carless (2006) reported frustration with a tutor's vague comment to 'elaborate more'. Another issue is one of the timeliness of feedback – how quickly the tutor is able to return students' work. Many studies indicate that tutors and students alike believe that quicker turnaround time is better (Williams and Kane, 2009; Robinson, Pope and Holyoak, 2011; Beaumont, O'Doherty and Shannon, 2011; Barker and Pinard, 2014; Mulliner and Tucker, 2017). (However, it is interesting that despite this emphasis on the importance of timeliness of feedback, which undoubtedly appeals to common sense, I was unable to find any studies which sought to quantify the impact of timeliness of feedback on learner uptake.) In Chanock (2000:104) undergraduates in L1 content courses (n=101) from various departments completed an anonymous voluntary four-question survey:

*Tutors often write on essays they are marking, something like: Too much description (/narration/ retelling); not enough analysis.*

1. *What do you think is the difference?*
2. *Have you ever had a comment like this in high school or university?*
3. *If so, was it clear to you, from this comment, how you could have made the essay better?*
4. *If you had such a comment, how did you feel about it?*

The results (from approximately 80 respondents) tended to demonstrate a lack of understanding about the provided example tutor comment given in the survey, as approximately half of the participants who reported having received a similar comment indicated that they did not understand what to do with it. It seems essential, then, that teachers ensure that students are able to understand the feedback that has been provided; this issue is addressed next.

#### *2.4.4.2 Confirmation that CF has been understood*

In the case of corrective feedback, a lack of understanding of CF could prevent learners from gaining an understanding the rules governing grammatical features, perhaps increasing the likelihood of reoccurrences of the error, possibly causing frustration to the learner and delaying L2 development. One way of appraising learner understanding is by asking learners to explain the rules governing the error in their own way; theoretical and empirical support for this will be provided in this section.

Before entering into a discussion, however, it is useful to clarify some terms. Consistent with the discussion in Section 2.3, *metalinguistic terminology* refers to knowledge of technical terms; participants may or may not use technical terms to express their metalinguistic knowledge. A learner's oral or written expression of their metalinguistic knowledge can be called a *metalinguistic reflection* (Simard, French and Fortier, 2007). The similarity of *metalinguistic reflection* to *metalinguistic explanation*, which was described in Section 2.4.2.2 as a form of CF, could cause confusion, but the two are not the same. A review of research suggests that *metalinguistic explanation* is the preferred term when referring to feedback given <u>to</u> a learner, and not when referring to the case when learners themselves are expressing their metalinguistic knowledge.

Confusingly, in addition to *metalinguistic reflection*, a number of other terms are also used. For example, the act of reflecting on language has also been described as *metatalk* (Swain, 1998), but problematically appears to exclude written communication. Swain (2006), seeking to clarify the issue and broaden the application beyond spoken reflection, introduces the term *languaging,* defined as 'the process of making meaning and shaping knowledge and experience through language' (Swain, 2006:98). The term has become widely accepted in the field (Knouzi et al., 2010; Suzuki, 2012; Ishikawa and Suzuki, 2016; Moradian, Miri and Hossein Nasab, 2017). Chi et al (1994:472) point out that an advantage of having learners express their knowledge is that it 'gives rise to multiple opportunities to see conflicts between one's evolving mental structure and the [target level of understanding]'. This fits well with the earlier discussion on noticing and feedback.

The idea of languaging exists in, and seems to have originated from, literature in the field of cognitive psychology, where it is called *self-explaining* (Chi et al., 1994). *Languaging* is the term applied in this study in order to maintain consistency with the field of applied linguistics, where this tends to be the preferred term. As can be seen in the summary of the discussion provided in Table 2-5, *languaging* tends to refer to a process, whereas *metalinguistic reflection* tends to be the product of that process.

**Table 2-5 Summary of terms and explanation of usage**

| Term | Notes |
|---|---|
| metatalk | Not commonly used in modern literature; problematic because it seems to exclude written reflection |
| self-explaining | Less commonly used in SLA literature, though occasionally does appear; comes from cognitive psychology |
| languaging | The process of a learner explaining linguistic notions |
| metalinguistic explanation | Most commonly refers to linguistic explanations given to a learner in the context of feedback provision (cf. Section 2.4.2.2) |
| metalinguistic reflection | A learner's self-produced linguistic explanation about features of their production |

Swain et al. (2009:5) see the act of languaging as beneficial to learning as it first 'transforms inner thoughts to external knowing … [and then] … transforms external knowing into internal cognitive activity' This suggests that one explanation for the hypothesised effect of languaging might be found in information processing (Section 2.3.1.2), as learners are encouraged to analyse the information, resulting in greater learning.

Another way languaging can assist learning is by making it clear to the learner where the boundaries of their knowledge currently are. A language learner in Knouzi et al. (2010) began the languaging process by expressing her understanding of a linguistic concept, but as she explained it to herself, she realised her understanding was incomplete. This realisation prompts her to consider the problem again, working independently toward a fuller understanding. Thus, it would seem that using CF with languaging could contribute to L2 development. Swain and Lapkin (2002) investigate languaging with two French L2 learners in grade 7 in a French immersion school in Canada. Participants first received instruction on the targeted grammatical feature, pronominal French verbs. Then they spent 30 minutes collaboratively writing a paragraph based on pictures they were given (pre-test; collaborative). In the next session, participants were given feedback (a reformulation) and their original work and asked to discuss together differences they noticed between the two; this was

videotaped. For the post-test, the researchers then gave each of the participants a copy of their original work and asked them to revise and rewrite it (post-test; individual). Of 47 issues the pair had discussed, the participants both individually corrected approximately 78%, which the authors see as suggesting that the process of drafting, noticing, and languaging may be able to contribute to learners' L2 development.

Despite a possible positive effect for oral languaging being indicated here and in other studies (Storch, 2008; Knouzi et al., 2010; Swain et al., 2009), larger-scale, quantitative empirical work on the topic was not found. This might be an effect of the procedural and resource demands of languaging; engaging large numbers of participants in one-to-one or pair (or even small-group) languaging could require a great deal of time and money (e.g. for transcription, etc.). However, there are other ways of investigating languaging. While languaging can be done orally with an interlocutor present, as here, research suggests that languaging can also be effective when learners write metalinguistic reflections, called *written languaging* (Suzuki, 2012; Ishikawa, 2013). Ishikawa and Suzuki (2016:98) define written languaging as 'learners' self-explaining of the target grammar rule in writing'. As discussed in Section 2.2, this could have the advantage of being a slower process which allows learners to consider their metalinguistic reflections, possibly allowing deeper processing, whereas speech, as in oral languaging, is an online activity, requiring learners to think and respond quickly. When teachers elicit metalinguistic reflections from students, it can be said that the teachers are inviting students to participate in languaging, be it spoken, as above or written, as will be discussed next.

Ishikawa (2013) examined how written languaging mediates longer-term learning using a pre-test/post-test/delayed post-test design. The participants were Japanese L1 EFL university learners (n=14; gender ratio not noted), divided into treatment and control groups. The focus of the study was tense consistency. For the treatment, participants were asked to write L1-L2 translations for four sentences, and then hide their translations and examine model answers; only the treatment group made notes about anything they noticed in the model answers. No revisions were made; the final step was the immediate post-test, completed without looking at notes or other written work from the treatment. Four weeks later, the delayed post-test was administered. The results showed that both treatment and control group improved from pre-test to post-test, but were not significantly different from each other. The treatment group maintained somewhat higher scores on the delayed post-test compared to the control group, but again, the difference between groups was not significant. It is likely that the small number of participants contributed to the difficulty attaining

statistical significance. It also seems possible that the treatment task, translating four sentences, was not enough to prompt the cognition that the researcher wished to see. More sentences could have provided more examples for students to interact with; with only four examples, there was little room for comparison and questioning.

Using what could be argued to be a more ecologically valid design, Suzuki's (2012) study also examined effects of written languaging with Japanese EFL learners at a university in Japan (n=24; 62.5% female; intermediate proficiency as judged by researcher). The study took place during two lessons spread over two weeks; in the first lesson, students were given 30 minutes to write an essay; in the second lesson, they received unfocused direct WCF, completed a languaging task and a background questionnaire, and revised their essays. The languaging task asked students to write in L1 why indicated linguistic errors had been marked as incorrect. The students' explanations of their errors were coded *Lexis*, G*rammar*, *Don't know*, and *Other,* with the last category being excluded from analysis, resulting in 522 items. Sixty-two percent of corrections were for grammar, of which 91% were successfully revised; 26% of corrections were for lexis, of which 96% were successfully revised; and the remaining 12% were from the *Don't know* category, of which 73% were successfully revised. Note the high successful correction rate for grammar; this might seem to contradict findings by Park (2013) and Gass et al.(2003), discussed previously, both of which reported a tendency for learners to focus on lexis more than syntax. In those studies, however, receptive rather than productive knowledge was tapped. It is possible that as students in Suzuki (2012) produced their writing, they became aware of gaps in their knowledge, which then might have primed them for noticing received feedback; one might imagine learners, upon seeing direct correction on forms they had struggled to produce, thinking, 'Ah, so that's how we say that!' Swain's output hypothesis supports this, explaining that learners' production triggers *noticing the hole,* which then allows learners to see holes in their linguistic knowledge that they may then consciously work to fill (Swain and Lapkin, 1995); and that then being presented with the target form allows them to *notice the gap* between what they have produced and the target form. Another feasible explanation for the contradictory findings between Suzuki (2012) and Gass et al. (2003) is that since the direct corrections in Suzuki did not contain explanations, perhaps the learners were unable to intuit or deduce differences between their chosen lexis and the correction. For example, one student wrote 'What I know about history is *small*', which was given direct correction: ~~*small*~~ → *limited*. On the languaging task, the student explained this error by writing that *limited* sounds better, but this does not indicate understanding, since *Because X sounds better than Y* could be used to describe most lexical problems. Finally, it seems unfair to have asked students to guess about reasons for unfocused lexical corrections since

they might lack knowledge upon which to base an answer unless the lexical item had been studied in class.

Ishikawa and Suzuki (2016) compared the effect of written languaging to that of other grammar practice activities and to a control group with Japanese L1 university students (n=40; 31 male; ages 18-20) in Japan. The targeted feature was 'the present counterfactual conditional (e.g., "If I had more time, I would travel around the world")' (p. 99). The pre-tests, post-tests, and one-week delayed post-tests were almost identical in design, with only minor differences (e.g. vocabulary) distinguishing them:

- Recognition test: fifteen multiple-choice questions testing participants' ability to identify English L2 sentences containing correct grammar (ten of these focused on the targeted structure, and five were distractors)
- Production test: seven Japanese L1 sentences for translation (five for the targeted structure; two were distractors)

Treatment group participants were given three minutes to read an explanation of the targeted structure; the control group did nothing. Then the explanations were collected and the written languaging group were asked to write their understanding of the rules governing this structure while the grammar practice group completed a grammar worksheet; again, the control group did nothing. Scores of the pre-, post-, and delayed post-tests were statistically tested to see the results. While an ANOVA revealed that all three groups had similar pre-test scores, there were statistically significant differences only for the Recognition test ($F(2, 38) = 6.84$, $p < 0.01$) between the control and the two treatment groups, but not between the treatment groups. A similar pattern was seen for the delayed post-test, with a separate ANOVA revealing significant differences between the control group and the two treatment groups (but not between treatment groups) on the Recognition test; only the written languaging treatment group is reported to have reached statistical significance on the Production test, but the results are, curiously, not provided. No effect sizes are provided, and it is unclear why three separate ANOVAs were conducted rather than one mixed ANOVA; regardless, with only 40 participants, who were divided between three groups, the decision to conduct statistical tests to measure differences between groups could be problematic (Dörnyei, 2007:100).

Having provided feedback and appraised learner understanding through metalinguistic reflections, tutors can require students to revise their work, incorporating the feedback, as this has been implicated in research as an influencing factor for CF effectiveness. One explanation for this is that, from a cognitive theoretical perspective, in the process of writing the revision, learners may attend

to their errors as they compare their original grammatical forms with the ones provided in the feedback, and that this noticing may result in improved accuracy in later drafts; from a human information processing perspective, this might relate to Craik and Lockhart's (1972) Type II processing, where learners move beyond trying merely to memorise the information, and begin analysing the feedback more deeply as they incorporate it into their original work. Chandler's (2003) study examined the effect of revisions on the effectiveness of indirect feedback provided to ESL students (N=31; various L1s; high-intermediate or advanced proficiency; all East Asian) at a conservatory in the USA. She required the treatment group to write revisions of their work, while the control group did not. After five writing assignments, each an autobiographical work, and corresponding iterations of feedback (unfocused indirect CF), the treatment group produced significantly ($p$<.001) fewer errors per 100 words of writing than on the first assignment while the numbers of errors per 100 words in the control group remained similar to their first assignment.

From these results, Chandler concludes that not requiring revisions is 'equivalent to giving no error feedback' (ibid.:280). Similar conclusions have been reached by other researchers (Shintani, Ellis and Suzuki, 2014). Furthermore, it seems it is not enough merely to recommend students do extra work that could benefit them. First-year university students (n=539) in L1 content courses in Fisher, Cavanagh, and Bowles' (2011) study were given the option of submitting a draft for feedback, but less than 20% of the students (n=106) accepted this offer. Such cases prompt O'Donovan, Rust, and Price (2016:942) to conclude that 'without encouragement or compulsion, a significant number of students may *not* take up drafting opportunities' (emphasis original). If students demonstrate reluctance to revise work before submitting it for a grade, as in O'Donovan et al., it could be expected that they may be even more reluctant to spend time revising their work after submission and feedback, making the *requirement* element of revisions an important to student learning.

#### *2.4.4.3 Facilitating learner use of CF*

Initially, learners' ability to use CF might seem better suited to being discussed in the section on learner-dependent factors, but as suggested in the discussion above, some learners may not independently take steps to further their learning, meaning that teachers might need to provide the necessary *encouragement or compulsion* (Fisher, Cavanagh and Bowles, 2011; O'Donovan, Rust and Price, 2016). However, even this may be insufficient. For example, having provided learners with CF and required them to write revisions, some learners may carry this out mechanically, failing to fully engage with the learning opportunity. Dörnyei and Kormos (2000) propose that task engagement is so important that it might even be considered a prerequisite for L2 development. They argue that 'if

students are not actively involved in the instructional tasks and do not produce a certain amount of language output, the tasks are unlikely to be effective'. The researchers investigated various motivational aspects of task engagement with Hungarian L1 English learners (ages 16-17; intermediate proficiency), measuring *engagement* as the number of words spoken by learners and the number of turns taken by learners. The task required participants to work individually to rank given items in a hypothetical situation, and then to work in pairs to compromise and agree on their rankings; this pairwork was audio-recorded and transcribed. Correlational tests performed on the engagement measures and motivational measures show, inter alia, a significant correlation between both engagement measures (words and turns) and attitudes toward the English course in general ($p<.05$); and between both engagement measures and attitudes toward the specific task ($p<.05$). The results suggest the possibility that learners can enter a given class with attitudes about the class that can influence how much they engage in class activities, which in turn may influence their learning. The results of the study also seem to suggest that even if a learner has a generally positive attitude about the class, a negative view of a task could influence the learner's engagement with class activities, which in turn may influence their learning.

Linking the findings of Dörnyei and Kormos (2000) to CF, it would seem that if learners are unable to use CF for some reason, then perhaps there is something teachers could do to help. The question is how to accomplish this so that it is most useful for the students. CF is 'at one and the same time … both a social and cognitive event' (Ellis, 2010:346). While the cognitive frameworks discussed provide a rationale for why and how corrective feedback could contribute to L2 development, sociocultural theory (SCT) 'offers a framework through which cognition can be investigated systematically without isolating it from social context or human agency' (Thorne, 2005:393). That is, SCT provides a way of thinking about the process of providing feedback, and can be viewed as providing a framework through which cognitive skills being guided by others can be investigated. As discussed in Section 2.3.1, information processing theory proposes that deeper analysis of information can result in increased learning. The next sections will discuss how the SCT concept of scaffolding can facilitate learners' use of CF by prompting analysis of it.

#### 2.4.4.3.1 Scaffolding

There are two sections dedicated to scaffolding. This section will discuss scaffolding as it is often understood in educational contexts, with the tutor supporting the learner. Section 2.4.4.3.2 will then look at self-scaffolding, which is another way scaffolding has been conceptualised in the literature.

As discussed in Section 2.3.1.1, while declarative knowledge can be, though is not necessarily, acquired rather quickly, 'procedural learning occurs only in executing a skill; one learns by doing' (p348). Anderson (1983) explains, first, the learner acquires declarative knowledge; then the learner begins perfecting performance through discovering and eliminating errors; once this production has been formed, the work of increasing speed and, through hypothesis testing, accuracy begins (p. 350ff). The disadvantage of this process, Anderson notes, is the time and effort needed. Therefore, it is efficient to move from purposefully interpreting declarative knowledge to '[applying] the knowledge directly' (p. 379). Moving from purposefully interpreting declarative knowledge to applying it, though, can be seen as presenting a paradox, as learners are expected to perform skills they have yet to acquire (Griggs, Carol and Bange, 2002). Though the theory of skill acquisition suggests that this happens through practice, resulting in *proceduralisation* (Section 2.3.1.1), some see the *practice → proceduralisation* equation as being too simplistic. Bange, Carol, and Griggs (2005:34) argue that the learner should not be expected to accomplish this independently, and that learners need to be supported in their practice as they move toward proceduralisation. Lyster and Sato (2013) argue that this suggests a specific role for teachers to provide the scaffolding that will allow learners to develop their L2 skills. Explaining the benefit of having support for CF, a language learner in Swain and Lapkin (2002:299) said, '''OK, so they changed something [on my paper]. But you don't know why they changed it. So that's why it's better to have the person like, talking to you about the corrections that they've made'''.

One way of providing such support, Bange, et al. (2005:34) point out, is through *scaffolding*, a concept credited to Wood, Bruner, and Ross (1976:90), who define it as a 'process that enables [a] novice to solve a problem, carry out a task or achieve a goal which would be beyond his [*sic*] unassisted efforts'. They identify six features which characterise scaffolding (ibid.:98), provided below, with relevant notes from other research as needed:

1. Recruitment – Focusing the learner's attention on the task
2. Reduction in degrees of freedom – Simplifying the task
3. Direction maintenance – Keeping the learner focused on the goal
4. Marking critical features – Accentuating 'features and discrepancies between what has been produced and the ideal solution' (Donato, 1994:41)
5. Frustration control – Preventing the learner from experiencing excessive amounts of frustration
6. Demonstration – Modelling the task

Stone (1998) argues that that misconceptions of scaffolding lead to less effective practical use of the concept. He explains that some of the misunderstanding originates with two different forms of speech for the term: noun versus verb. The author insists that scaffoldings (noun) are important to

the learning process, but 'what is most crucial is the process by which these entities [scaffoldings (noun)] are used to foster new understandings … the core of the scaffolding metaphor rests squarely on viewing it as a process' (ibid.:412). Lepper, Drake, and O'Donnell-Johnson (1997), on the other hand, seem to accept the idea of scaffolding-as-noun, and instead blame the terminology for misconceptions that have developed. They write that the idea of scaffolding 'carries the inappropriate connotation that the learner (like the painter or the window-washer) will return to ground level when the scaffolding is removed' (ibid.:110). They offer an alternative metaphor of scaffolding used in the construction of arches or tunnels, which stand independently after the scaffolding removed. Puntambekar & Hubscher (2005) echo the criticism offered by Lepper, et al. (1997), pointing out how the concept of scaffolding was originally to include a plan for fading out the support, but modern users tend to maintain the support. Puntambekar & Hubscher (2005) also argue that two more crucial elements have been lost in recent years: a) that scaffolding was intended to contain an element of ongoing assessment of individual learners, and b) that the support being offered was to be tailored to the needs of the individuals, as the assessments dictated; all three points, they argue, are necessary for successful expert scaffolding.

Much of the empirical work on scaffolding has been conducted without statistical analyses (Ko, Schallert and Walters, 2003; L. Lee, 2008; Thompson, 2009; Kayi-Aydar, 2013); notably, though, all of these studies mention the importance of learner engagement for the success of the scaffolding. Though not focusing on language learning, Finn and Metcalfe (2010) investigated the comparative effects of a) scaffolding (provision of incremental hints) with those of three other types of feedback: b) direct feedback; c) answer-until-correct multiple choice feedback; and d) minimal feedback (informed whether the answer is correct or not) on a trivia-type computer game; all questions had one-word answers. Undergraduate students (n=18) volunteered for the study in exchange for cash or course credit. Encouraged to guess if they did not know the answer, participants answered questions (pre-test; from a pool of 191 questions; difficulty measures not reported) until they had 40 instances of incorrect answers and corresponding feedback (treatment).

Those 40 questions, which varied according to participant, became the post-test, which was administered to all participants immediately for half of the questions, and 1-2 days later for the other half of the questions. Results of unnamed statistical tests of significance indicate a significant interaction between group and time, $F(3,51) = 4.85$, $MS_e = 0.03$, $p < .05$, $\eta_p^2 = .22$; post-hoc tests revealed no significant difference between the scaffolded, direct, and answer-until-correct feedback groups at the time of the immediate posttest, but the one- to two-day delayed post-test did show

significant differences between the scaffolded group and the standard feedback group, and between the scaffolded group and the answer-until-correct group ($p<.05$). The authors interpret their findings to suggest that for longer-term memory, scaffolded feedback is superior to other forms of feedback. With 18 participants divided between four groups, however, the decision to conduct statistical tests seems questionable. It seems that deeper investigations with sufficiently high numbers of participants need to be conducted in order to understand both how effective scaffolding is in a language learning environment and how learners respond to it.

#### 2.4.4.3.2 Self-scaffolding

The previous section discussed scaffolding by an expert; another way of inviting learners to focus on targeted grammatical forms is through *self-scaffolding,* a concept commonly attributed to Bickhard (1992a; 1992b). If scaffolding is, as above, assistance by an expert to a novice, then the term *self-scaffolding* might seem to be an oxymoron: '[a] person cannot provide to him or her self-knowledge [sic] that is otherwise not available' (Bickhard, 2005:170). In answer to this, Bickhard introduces the concept of *functional scaffolding*, which conceptualises *scaffolding* as a series of resources ranging from more to less accessible (from the perspective of the learner); that is, in earlier stages of learning, learners might benefit most from resources provided to them by an expert other; as learning progresses, these expert-provided resources could be replaced with ones sought out by the learner, keeping the complexity of the task challenging, but within reach of the learner. That is, at one end of the continuum there is much support, with the opposite end offering less support. Bickhard sees self-scaffolding as contributing to learning to learn, or 'knowledge of how to approach problems' (ibid.:171); an example of this could be encouraging the learner to find answers to their questions by reviewing their class notes and textbooks rather than viewing the teacher as the sole source of information.

Bickhard's conceptualisation of functional scaffolding supports the idea of self-scaffolding by focusing on the use of resources available to the learner rather than on the use of an expert. Bickhard writes that scaffolding, including self-scaffolding, can be accomplished in several ways, such as 'by choosing simple cases to work on first, by moving to idealizations, by breaking down into subproblems, by making use of resources that are currently available but may not always be available, and so on' (ibid.:170). Viewing the provision of feedback through the lens of self-scaffolding, it seems that the process would be: a) marking the errors of the targeted grammatical forms on the students' work, thereby breaking the work to be done (corrections) down into subproblems and identifying the cases for the learner to focus on, and b) requiring learners to refer

to previously provided material, drawing upon what they had been taught, thereby requiring them to '[make] use of resources that are currently available' (p170).

While self-scaffolding is, as mentioned, attributed to Bickhard, he is not alone in his argument for the concept; others argue a number of reasons to support self-scaffolding. For example, some find the traditional concept to be overly prescriptive (Littleton, 2013) or too linear (Mascolo, 2005); others argue that traditional conceptualisations are one-sided, failing to acknowledge agency of the learner (Granott, 2005). With this support in the literature, the concept of self-scaffolding has recently been applied to language learning. As was the case for scaffolding in the previous section, studies on self-scaffolding also tend to be smaller scale and/ or without statistical analyses. Lapkin, Swain, and Knouzi (2009) investigated a form of self-scaffolding with English L1 (or English-dominant) learners of French (n=9) in an intermediate French course in a university in Canada. Instead of seeing what errors participants made and then asking them to explain those, *self-scaffolding* in this study was proactive; that is, learners read through explanations of aspects of the targeted feature (voice), and participated in languaging to reach a fuller understanding of the grammar explanations on the cards. To test whether languaging was effective, participants were asked to read a prepared text and to explain the bolded grammatical features in it. Then they were presented with explanations of aspects of the grammatical features on individual cards and asked to read them and carry out languaging about the explanations. Participants then took the immediate post-test, which again contained bolded terms about which participants were to engage in languaging in their L1. To test for longer-term effects, a one-week delayed post-test was administered; here, participants were asked to complete a cloze test of the text they had read in earlier stages of the study; how comparable this delayed post-test is considered to be to the pre-test and post-test is not discussed. The researchers analysed the participants' languaging, finding significant correlations between the amount of languaging and the immediate post-test, but not between languaging and the delayed post-test ($\tau = .572$; $p<.05$); it seems worth noting that with only nine participants, the choice to statistically analyse the data might be considered questionable.

Knouzi et al. (2010) used the same data from the study above to analyse the two participants who produced the highest and lowest amount of languaging, finding that the two participants were not equal in their demonstrated understanding of the task. The lower languager, for example, was reported to use simpler definitions and explanations; she also seemed unbothered by concepts she clearly did not fully grasp, for example, saying 'I don't really know' but quickly moving on instead of analysing the concept. The authors refer to this participant as a low languager, but they do not

consider possible reasons why she might have performed as such. Even though the participants were drawn from a course for intermediate French learners, no proficiency test was administered, and it is unclear whether this course had prerequisites. It is possible that the low languager was merely of a lower proficiency, and that the task was too difficult for her; the low languager was also from an immersion background rather than from a 'Core French' background, which seems to have a more systematic focus on grammar. Given the information in Section 2.4.3.1 about the impact of learners' educational and experiential histories on their responses to CF, it might be unsurprising that the low languager was less likely to analyse grammar and displayed greater comfort with the state of not knowing the rules. Further research is necessary to tease apart the influencing factors; there is also a need for research to demonstrate the effectiveness of scaffolding with statistical analyses.

## 2.5 Chapter Summary

The chapter began with a discussion of how writing can benefit learners, particularly second language learners. Writing is a mode of output that has permanence, providing learners with the opportunity to work slowly and to reflect on and revise their output (Section 2.2). Discussed next were the cognitive theoretical frameworks that inform the understanding of language learning as discussed in this thesis, with the interaction hypothesis drawing from other cognitive theories to explain L2 grammar development (Section 2.3.1). Having established this basis, the chapter examined the value of explicit knowledge, finding two main ways it can benefit learning. Explicit knowledge is hypothesised to be useful to language learning in that it might, at some point, become implicit knowledge, but even if the knowledge does not become implicit, learners can benefit from using explicit knowledge to monitor their writing, checking for errors and correcting their work (Section 2.3.2.1). Having discussed the value of explicit knowledge, the discussion examined how learners can gain and develop explicit knowledge through instruction.

Various ways of providing instruction have been identified, and they may each be best suited to certain groups of learners. For example, there is much support for focusing learners' attention on forms, and, for lower-proficient learners, also for isolating the instruction of the targeted linguistic forms (Section 2.3.2.2). As learners gain explicit, declarative knowledge of linguistic forms, skill acquisition theory suggests that by practicing these forms, learners can begin proceduralising that knowledge (Section 2.3.1.1). However, as learners are prone to make errors when using the targeted linguistic forms, it may be helpful to draw their attention to the errors so that they can notice the gap between their interlanguage and the targeted form, and try to use the correct form in later writing (Section 2.3.1.3).

CF, the focus of Section 2.4.2, is a common way of drawing learners' attention to their errors. There are different ways of providing CF on learners' writing, such as direct CF, indirect CF, or metalinguistic CF. Indirect CF was seen to be problematic for some learners, such as those with lower L2 proficiency, who might struggle to understand the problem. Direct CF provides the learner with the correct form, thereby removing ambiguity, but it can fail to have an enduring effect, possibly because, as information processing theory suggests, it does not encourage learners to engage in deeper analysis of the issue. Metalinguistic CF, by providing more information about the error, might encourage the learner to analyse the information, resulting in greater learning. However, while metalinguistic CF might be argued that it is theoretically superior to other CF types for lower-proficient learners, as with all forms of CF, if learners are unable to use it, it can still fail to be effective.

There are a number of factors suggested by research as possibly influencing learners' ability to use CF. Some of these, as in Section 2.4.3, are related to the learner, such as a) their learning context, including prior learning experiences; b) their interests and preferences and the way these interests may lead students' studies and foci; c) their self-efficacy in their ability to use CF; and d) their language aptitude, specifically, grammatical sensitivity. Despite individual learner differences, there are steps teachers can take to increase CF effectiveness (Section 2.4.4). For example, teachers can ensure that the CF is provided in a way that is clear to the learners, and they can take steps to ensure that the CF is understood, such as by asking students to produce metalinguistic reflections (Section 2.3.2.1). Since learners might be unable to make independent use of CF, teachers can facilitate students' use of CF, thereby encouraging the deeper analysis hypothesised by information processing as contributing to learning, helping learners benefit from the CF, but this has largely not been addressed in CF studies.

The current study, explained in the next chapter, seeks to investigate whether teachers can facilitate learners' use of CF in a way that results in learners' increased ability to use CF to improve their grammatical accuracy in writing.

# 3 Methodology

## 3.1 Introduction

This chapter will explain the design of the current study and of the instruments and procedures applied in the research. The chapter begins by briefly explaining the study's aims, followed by a description of the methodological orientation of the study in Section 3.3. The chapter then provides a description of the research design, including the research questions. Following this, the chapter provides a brief discussion of the piloting of the study (Section 3.5). This information about the piloting is provided early in the chapter because it will be referred to throughout the description of the current study, as the piloting procedures helped to shape the current study. Next, the participants and setting of the current study are described. Following this, the current study is described in Section 3.6. Following the discussion on terminology in Section 2.3, as the current study focuses on learners during a short period of time, what happens during a particular stage is more useful than a broad view of language learning. Therefore, *L2 development* will be used in discussions.

## 3.2 Aims of research

The over-arching aim of this investigation, in which first language, culture (broadly speaking), and disciplinary orientation are controlled, is to contribute to the body of knowledge on written corrective feedback in the growing numbers of universities worldwide where non-English specialists are required to learn and write in English. This study is unique in its investigation of the effectiveness of the different modes of CF facilitation for developing the explicit knowledge of grammatical features in the written English of language learners who are Korean L1 STEM majors. It has three specific aims which contribute to the overarching aim of the study. First, it aims to contrast the development of explicit L2 knowledge in two treatment conditions where CF is facilitated and compare these results to those of a comparison condition (direct written CF). Secondly, the study investigates whether language analytic ability and self-efficacy play a role in learners' improvement in grammatical accuracy, as has been proposed in the literature. Thirdly, the study investigates learner and teacher experiences of the three conditions, on the basis that perceptions of effectiveness, efficiency, and ease of use will have an impact on uptake by learners and teachers. Theoretically, the study contributes mainly to the development of the interaction hypothesis, which, as was seen in Section 2.3.1.3, draws from skill acquisition theory and information processing. It is hoped that the study will reveal whether explicit L2 knowledge can be developed by facilitating learners' use of CF, as evidenced by increased accuracy of the targeted forms in written output, and whether these effects are greater than those of a group not receiving facilitated CF. Pedagogically, it

is hoped that the results of the study will reveal whether self-scaffolding via metalinguistic reflections with worksheets might be recommended to teachers as an effective alternative to time-consuming one-to-one conferences, and whether self-scaffolding via metalinguistic reflections with worksheets is more effective than just writing corrections on students' work in improving performance in developing self-efficacy with regard to writing in L2 English.

## 3.3 Methodological orientation

Philosophical beliefs, whether understood and made explicit or not, influence how people think and the decisions they make. These beliefs can explain 'how some people can come to think in certain ways, which are bound by certain cultural and social norms and parameters, for example those established by disciplines in academia' (Grix, 2002:178). For researchers, these views and beliefs determine *which* questions are asked, *how* they are asked, and *how* the answers are found (ibid.). It is usually considered desirable for researchers to be as unbiased as possible in their work, but often, despite the best intentions, some amount of bias lingers as a result of the philosophical beliefs held by the researcher. One way of reducing the effects of this remaining bias is to highlight possible sources of bias so that the bias is more visible. To this end, this section will offer a discussion about the relationship between the theory of the study and my views as the researcher.

In the 1980s and 1990s, social scientists who used either quantitative or qualitative methods of analysis argued for the superiority of their respective research approaches based on the belief that positivist and interpretivist epistemological paradigms could not be held simultaneously, a debate which Howe (1988) called the *paradigm wars*. While some (e.g. Biesta, 2010:98) still classify *data* as quantitative ('expressed in numbers'), or qualitative ('usually expressed in text'), it seems more useful to accept Bryman's (2012:714f) position that quantitative and qualitative methods are *approaches to data analysis*, because this explanation makes room for the possibility, for example, of quantifying textual data or describing numerical data; and can help avoid unnecessarily contributing to paradigm wars. Modern researchers tend to agree that mixing quantitative and qualitative methodologies is unproblematic, or even unavoidable. They argue, for example, that

- education is a social science peg that cannot be forced into a natural science hole (Gage, 1989)
- there is such diversity in the field that it cannot be limited to merely two views (Hammersley, 2006)

- neither quantitative nor qualitative research methods are pure, that tensions exist within each, and that because of these tensions and differences, quantitative and qualitative methods are sometimes not very different from each other (Datta, 2006)
- social scientists should choose the best tools from all that are available to them to accomplish their goals (Oakley, 2006)

As a result, some social science researchers might be willing to reject positivism. Even so, they may also feel uncomfortable with interpretivism, from which view some might reject concepts of generalisation, prediction and control in research. A third solution that might offer balance is *pragmatism*. Current understanding of pragmatism has tended to embrace Dewey's interpretation of the philosophy, built on earlier conceptualisations of pragmatism by Peirce and James (Talisse and Aikin, 2011). Dewey wished to 'forward the emancipation of philosophy from too intimate and exclusive attachment to traditional problems. It…raises a question *as to the genuineness, under the present conditions of science and social life, of the problems*' (ibid:.110; emphasis original). That is, Dewey wished to bring philosophical thinking from its traditional place in the realm of the metaphysical down to the practical reality. This conceptualisation sees knowledge as being a social entity which is mediated by signs, and conceptualises the scientists as agents who are striving to solve real-world problems (ibid.), and has 'its emphasis on human beings as agents and their practical relations to the world' (Delanty and Strydom, 2003: 277f). Dewey (1917; reprinted in Talisse & Aikin, 2011) insisted that philosophy must address current-day problems and not remain, as he stated, 'distant and esoteric'.

Especially if the methodology develops from a pragmatic view, mixed methods research (MMR) has been proposed as a research method capable of rising above the quantitative-qualitative dispute (Onwuegbuzie and Leech, 2005). To synthesise and update existing definitions of MMR, Johnson, Onwuegbuzie, and Turner (2007) asked MMR researchers (n=31) to report their own working definitions of MMR. They analysed the nineteen responses they received and coded the contents of the definitions. They found five common components that researchers tended to include in their definitions: 1) That which is mixed: i.e. quantitative and qualitative; 2) The stage at which mixing occurs: when and where; 3) The breadth of the mixing on a continuum: from simply collecting both *quantitative and qualitative data*[2] to mixing at all stages; 4) The reason for mixing: e.g. breadth; or corroboration, such as triangulation; 5) The orientation of the research: deductive or inductive. The

[2] Note, though, the problem discussed previously with classifying data (rather than methods of analysis) as *quantitative* or *qualitative*.

resulting definition of Johnson et al.'s (2007:129) study is that MMR is

> the research paradigm that (a) partners with the philosophy of pragmatism [in its various interpretations]; (b) follows the logic of mixed methods research (including the logic of the fundamental principle and any other useful logics imported from qualitative or quantitative research that are helpful for producing defensible and usable research findings); (c) relies on qualitative and quantitative viewpoints, data collection, analysis, and inference techniques combined according to the logic of mixed methods research to address one's research question(s); and (d) is cognizant, appreciative, and inclusive of local and broader sociopolitical realities, resources, and needs.

The definition is broad and deep, having emerged from a synthesis of current expert-derived definitions, and it is relevant, reflecting perceptions of MMR which have evolved over time with research. It succeeds in a number of ways. It acknowledges the complexity of social science research by providing guidelines which preserve the centrality of the research question (c, d). It acknowledges the relevance of philosophy to research (a, b) while conceding that different research questions will require different techniques for obtaining answers (c). It neither restricts mixing nor creates an arbitrary minimum amount of mixing (b). Though the definition is quite lengthy, the complexity of MMR seems to demand it be so. It encourages a pragmatic view of context and flexibility in addressing research questions.

The current study applies MMR methodology, seeking to expand the depth of the quantitatively analysed data by qualitatively exploring student work and gaining insight into teachers' and students' thoughts on the interventions.

## 3.4 Research design

In order to achieve the aims of the study, a quasi-experimental, longitudinal, mixed methods design was implemented. The intervention itself is outlined briefly in Table 3-1, and a summary table of the whole study is provided at the end of Section 3.6.5 in Table 3-13. The study is situated within the writing to learn (language) dimension of research on writing (Manchón, 2011; Section 2.2), and the design is based on the literature on corrective feedback, on noticing, and on explicit grammar instruction. CF was defined in the previous chapter as *1)* someone's reaction or response *2)* to a learner's non-standard production *3)* for the purpose of increasing the linguistic accuracy of the production. The study uses cognitive theoretical concepts as a basis for understanding language development (Section 2.3.1), and sociocultural theoretical concepts as a basis for understanding how students engage with the interventions (Section 2.4.4.3). The accuracy of targeted features is quantitatively measured in samples of student writing throughout a university term, and follow-up interviews with student and teacher participants are used to gain insight into participant perceptions.

Menard (2002:35f) offers a few words of caution regarding the use of longitudinal designs: one is that time might alter how variables are defined and measured. This criticism is unlikely to apply to the variables in the current study because the length of the study was limited to fifteen weeks and the same measures were used at all time points. A second risk Menard (2002:38f) mentions is that of attrition. Since successful completion of the course in this study was a graduation requirement, and since the course lasted only one university term, it was judged that attrition was not likely to be a serious problem; and indeed, all but three study participants remained at the end of the fifteen weeks of the study.

Research questions 1 and 2 examine differential effects of facilitating feedback use on outcomes, with RQ1 (a and b) being the core of the study. RQs 1a and 1b investigate the effect of the interventions on outcomes; that is, the change in accuracy of targeted grammatical features in new writing. These RQs compare the results of the two interventions against each other and against those of a comparison group. RQ2 examines the effect of the interventions on student self-efficacy, attempting to examine whether a particular method of facilitating feedback use resulted in students feeling a greater sense of self-efficacy at the end of the term compared to when the interventions began.

RQs 3 and 4 examine mediating variables which might influence the outcomes. RQ3a examines the effects of language analytic ability and initial self-efficacy on change in accuracy of all targeted grammatical features. RQ3b was originally intended to offer a breakdown of all of the grammatical features, examining how the accuracy of each targeted grammatical feature changed over the course of the study. In the pilot studies (Section 3.4.1) participants had been required to use the targeted features in their writing, ensuring that all features were used, but unfortunately, the teacher-participants in the current study were uncomfortable with this requirement, worrying that it would cause the students to feel undue stress. Consequently, it was found that only articles were used frequently enough for analyses to be carried out. Therefore, RQs 3a and 3b offer an examination of the effects of language analytic ability and initial self-efficacy on change in accuracy of articles only. RQs 4a and 4b relate to the Reflective worksheet group only, examining whether the quality of languaging influenced outcomes, focusing both on all targeted grammatical features (RQ4a) and, for the reason explained above, on articles only (RQ4b).

RQ5 seeks to explore the students' (RQ5a) and teachers' (RQ5b) perceptions of the study, allowing for triangulation and for the interventions to be examined from the perspectives of both sets of

stakeholders. This is important because the study aims to contribute ecologically valid, pedagogically useful knowledge to the field (*cf*. Section 1.3). If students do not care for a CF intervention, this could result in lower levels of engagement with it; this lower engagement could impact on the effectiveness of the intervention (Storch, 2008). Likewise, if teachers find a CF intervention to be impracticable or otherwise unworthy of the investment in its implementation, then they may resist applying the idea to their classrooms (Donnell and Gettinger, 2015).

**Table 3-1 Intervention timetable**

| | Session 1 | Session 2 |
|---|---|---|
| Week 7 | *(general lessons, continued)* | First writing sample (**Pre-test**) |
| Week 8 | **1st iteration of intervention**: Feedback by group | Second writing task |
| Week 9 | **2nd iteration of intervention**: Feedback by group | Third writing sample (**Post-test**) |
| Week 14 | Fourth writing sample (**Delayed post-test**) | Post-study feedback |

### 3.4.1 Research questions

The research questions explained above can be summarised as follows:

**Differential effects of facilitating feedback use on outcomes**

1 What is the effect of feedback delivery mode on:
   a the ratio of all targeted errors per hundred words?
   b the ratio of correct article use to all obligatory article use?

2 What is the effect of feedback delivery mode on the development self-efficacy between initial and final measurements?

**Mediating variables which might influence outcomes**

3 What are the effects of language analytic ability and initial self-efficacy on:
   a the ratio of all targeted errors per hundred words?
   b the ratio of correct article use to all obligatory article use?

4 For the Reflective worksheet group, what is the effect of the quality of reflection in the worksheets on:
   a the ratio of all targeted errors per hundred words?
   b the ratio of correct article use to all obligatory article use?

**Participant thoughts and experiences**

5 Regarding the feedback delivery modes used in this study:

a what are students' perceptions of their experiences?
b what are teachers' perceptions of their experiences?

### 3.4.2 Operationalising facilitated feedback

Section 2.4.4.3 presented *scaffolding* and *self-scaffolding* as two ways of facilitating learners' use of feedback. This section will explain how scaffolding and self-scaffolding have been operationalised for the interventions in this study.

#### *3.4.2.1 Scaffolding in Conferences*

One way tutors can facilitate students' use of CF is by scaffolding students' metalinguistic reflections in one-to-one student-tutor conferences (cf. Section 2.4.4.3.1). Carnicelli (1980) went as far as to argue that the value of conferences outweighs that of classes, and that if time is limited, the time is better spent on conferences than classes. Research suggests that students tend to agree: 83% of students (undergraduate; L1 content) in Murtagh and Baker (2009) rated one-to-one conferences as an *effective* or *highly effective* feedback method. When asked to explain, students listed reasons, such as addressing problems, asking questions, and getting more information about written comments. They also cited affective benefits, such as increased confidence and self-esteem. Likewise, Klenowski (1995) indicates the power of emotion in feedback; he explains that conferences can help students feel more positive about their work, possibly resulting in higher self-efficacy. He wrote that 'Some of [the students] don't realise they have achieved anything, so it's not until discussion with you as a mentor that they start realising that: 'Yeah, OK, I've done that' (ibid.:153).

Research suggests ways teachers can increase the effectiveness of their conferences. Aljaafreh and Lantolf's (1994) Regulatory Scale moves through a series of steps designed to scaffold the learner as much as is necessary, though the authors themselves do not use the term *scaffolding*. The initial steps are preparatory: asking learners to self-correct their errors independently and to attend a conference with the tutor. Section 2.4.2.2 discussed how indirect feedback can be problematic for some populations, such as those with lower L2 proficiency, so this step might not be appropriate for all students. Next, Aljaafreh and Lantolf suggest that 'Tutor indicates that something may be wrong… [e.g.] "Is there anything wrong with this sentence?"' (ibid.:471). The learner is then given two more prompts (Steps 5 and 6) to help them locate the error being addressed, each one growing progressively more specific, before the tutor reveals the problem to the student, if necessary (Step 7). The learner is then given two opportunities to provide the correction (Steps 8 and 9). Steps 10-12 ask the tutor to provide the correct form, an explanation for it, and some examples.

Nassaji and Swain (2000) apply Aljaafreh and Lantolf's (1994) Regulatory Scale to their study of feedback for Korean L1 adult learners of English L2 (n=2; both female) in an ESL setting. The following is a transcript from one of the conferences in which the scale was applied (Nassaji and Swain, 2000:40f):

1 **T:** "Most of Korean have same name". So do you see anything wrong with this sentence?
2 **S:** uhm … I imagine that… uhm … but … in this position… I imagine that … I think it's ok [laughing].
3 **T:** It's ok?
4 **S:** Yes.
5 **T:** ah… what about this part. "most of Korean". Do you see … Do you see anything wrong with this phrase?
6 **S:** Most of a Korean … Korean [whispering] … mmm …
7 **T:** Anything wrong within this part [referring to the place of the error]?
8 **S:** Korean?
9 **T:** No.
10 **S:** No? Uhm …
11 **T:** Yes there is something wrong with this [referring to the word "Korean"].
12 **S:** Yes. I think … ah … Korea means Korean people, yes?
13 **T:** Ok.
14 **S:** Yes?
15 **T:** But grammatically, is there anything wrong?
16 **S:** Of the? No?
17 **T:** Yes, what?
18 **S:** Of the?
19 **T:** The, yeah.
20 **S:** Yes.

The twenty-line exchange addresses a single error – the missing *the* in *Most of Korean* – and this represents only the correction aspect of the scale, without the steps for checking comprehension. The tutor might then repeat the steps to address whether the writer had intended to mean *Koreans generally* (i.e. *Most Koreans*), and then go through all of the steps again to add the plural *-s* to *Korean*. All of this would still address only the first three words of the student's writing. Reading the transcript, it can be seen that the student does not see the error, confirming twice that she thinks the sentence is correct (Lines 2 and 4), and then guesses four times about what it is that the tutor might be referring to (Lines 6, 8, 12, 16). Though Nassaji and Swain (2000) do not mention whether the student in this case felt frustrated, it would be understandable if a student unfamiliar with the use of leading questions, such as these, might begin to feel like the tutor is indeed playing some kind of game, as discussed in the literature review. Therefore, the decision was made in the current study to apply a condensed version of this scale, limiting the amount of questioning in order to increase

efficiency and to prevent participants from feeling frustrated or feeling as though their teacher were playing a game with them.

The conference protocol follows Wood, et al.'s (1976) six features of scaffolding (Section 2.4.4.3.1) and is based on Aljaafreh and Lantolf's (1994) Regulatory Scale. Following the above discussion, teachers were instructed not to dwell on any single error, but to keep the conference moving forward. For each numbered error on the photocopy, the teacher was asked to use the following two-step contingent feedback protocol to help participants: 1) identify the error on their own paper, and 2) describe the relevant rule.

Step 1: Prompt the student to identify the error by pointing to the line containing the error and asking *Do you see a problem here?*

1a) If the student can identify the error, the teacher should move to Step 2.

1b) If the student is unable to identify the error, the teacher should identify it and then move to Step 2.

Step 2: Prompt the student to comment on why the correction was needed.

2a) If the student has demonstrated understanding of the rule, the correction is written on the student's paragraph, and the evaluation continues with the next error.

2b) If the student cannot demonstrate their understanding of the rule, the teacher first provides the rule and then writes the correction on the student's paragraph before moving on to the next problem area.

#### *3.4.2.2 Self-scaffolding with Reflective worksheets*

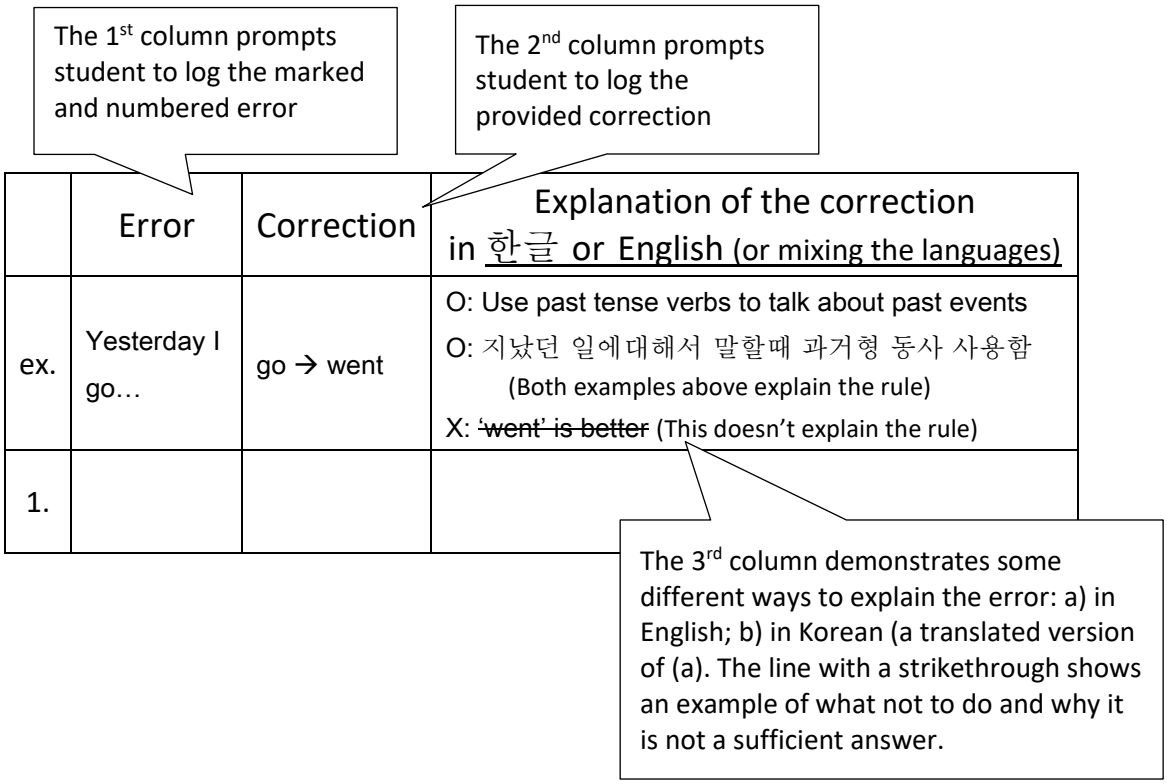
Another way tutors can facilitate students' use of CF is by encouraging the self-scaffolding of metalinguistic reflections via Reflective worksheets (Section 2.4.4.3.2). As the learner works independently, this is can be called *self-scaffolding*. Self-scaffolding in language learning literature tends to be conducted orally (cf. Section 2.4.4.3.2). However, given the practical and theoretical advantages of writing discussed in the literature review (e.g. writing being a slower process than online speech, thereby allowing deeper analysis of the content), it would appear that written self-scaffolding could offer similar or greater benefit to learners. This has been demonstrated in other areas of educational research. Toh et al. (2014) investigated participants' use of the scaffolding provided by a researcher-designed worksheet for solving mathematical problems. The participants were student teachers (n=55), and the worksheet was designed to 'explicitly [guide] the students to

go through the [process]… of problem solving' (p. 244). Though results are not measured statistically, the study reports that qualitative examination of participants' work demonstrated that participants were using the worksheet to scaffold their thinking. A separate study by the authors reported similar positive (but non-statistically tested) results with the worksheets with secondary school students (Dindyal et al., 2014), suggesting that self-scaffolding through worksheets can be appropriate for a range of ages. As literature on self-scaffolding through worksheets was scarce, the six features of scaffolding (Section 2.4.4.3.1) were applied in the development of the worksheets.

1) Recruitment – (Accomplished through the handing out of the worksheets)
2) Reduction in degrees of freedom – The learners should know where to focus their attention. Therefore, errors to be marked on the worksheet were numbered on the students' work to correspond to the numbers on the worksheets.
3) Direction maintenance – The worksheet task should be clear and unambiguous.
4) Marking critical features – Learners' attention should be drawn to the gap between their production and the target production
5) Frustration control – Completing the worksheet should be possible with learners' existing textbook and notes
6) Demonstration – An example demonstrates that both explanations and use of metalanguage terminology are appropriate, and that this can be accomplished in L1 or L2, or a combination of languages.

Figure 3-1 shows the first three rows of the worksheet: the column headings, the example, and the first space for logging an error.

**Figure 3-1 Reflective worksheet example**




In sum, *Conferences* can be viewed unproblematically either through the scaffolding lens provided by Wood, Bruner, and Ross (*experts guiding novices*), or through that provided by Bickhard (*resources on a continuum of more to less support*); that is, conferences are provided by experts, who offer contingent support to promote learners' understanding through interactive languaging. *Reflective worksheets* are best viewed through the lens of Bickhard's functional scaffolding, specifically as opportunities for self-scaffolding, as learners are invited to use tutor-provided direct CF and existing resources available to them to engage in written languaging to promote their own understanding. Table 3-2 provides a summary of the three treatments, along with an indication of how each treatment relates to the cognitive and sociocultural theoretical frameworks discussed.

**Table 3-2 Summary of intervention groups and corresponding theoretical frameworks**

| Group | Intervention | CF strategy taxonomy | SCT taxonomy |
|---|---|---|---|
| Conferencing group | Metalinguistic reflections elicited orally one to one | Explicit output-prompting CF | Expert-novice scaffolding |
| Reflective worksheet group | Metalinguistic reflections elicited in writing on reflective worksheets | Explicit output-prompting CF | Self-scaffolding |
| Comparison group | Direct WCF | Explicit input-providing CF | No contingent support |

## 3.5 Piloting the study

Before explaining the current study, the following sections will describe the two pilot studies that were conducted in preparation for the current study, highlighting which aspects of the piloting influenced the design of the current study; to maintain clarity, changes made to individual instruments are described alongside the relevant instruments in Section 3.6.3. The pilot studies were conducted during two successive terms in the year prior to the current study, one in each term, and each study made unique contributions to the current study. The two pilot studies involved the same course at the same institution with similar groups of students as the current study.

### 3.5.1 Pilot study I (PS-I)

The first pilot study, part of the author's MSc research (Boggs, 2013), was an intervention study comparing the effects of direct and indirect feedback in the first stage of the study, and then the effects of direct feedback to one-to-one conferences in the second stage. This informed the choice of interventions to apply in the current study, serving as a trial run of the comparison between one-to-one conferences and direct WCF and as an opportunity to refine the protocol for the conferences on the basis of literature in the field.

PS-I was carried out with the term cohort (n=90) just prior to the one participating in the current study. PS-I participants were divided into three groups: two treatment groups and one control group. Eight specific grammatical features were taught in classes, drawing students' attention to rules, common exceptions, and example sentences. Using a pre-test – post-test – delayed post-test design with four samples of new student writing, the pilot study first sought to compare the effects of indirect WCF to direct WCF. While the paragraph writing task (the same task as will be described in Section 3.6.5) was found to be appropriate (appropriate level of difficulty; not taking too little or too

much time), the students found it difficult and frustrating to independently use indirect WCF. The second part of the study then compared the effects of direct WCF to conferences, where the students had one-to-one sessions of four to five minutes with the teacher to discuss grammatical errors in their writing. Conference group participants demonstrated a statistically significant increase in grammatical accuracy over the control group. Another result that influenced the current study was an apparent threshold effect for a given type of feedback; the initial round of feedback failed to produce significant increases in grammatical accuracy, but there were significant increases after a second iteration of the same kind of feedback.

PS-I participants behaved quite shyly and seemed reticent in the conferences, which took place in a quiet room with only the teacher-researcher, and this informed the design of the current study, as will be explained in Section 3.6.3.1.5. Another change relating to apparent student shyness resulted from the piloting of the interviews. Though researchers have traditionally held such interviews in quiet rooms, it had seemed that PS-I participants were intimidated by being alone in a quiet space with the non-Korean teacher-researcher; anecdotally, discomfort is a common reaction in Korea when an individual Korean finds themselves alone with a non-Korean (*cf* Section 1.2.2). The need to put students at ease influenced the design of the interviews in the current study.

### 3.5.2 Pilot study II (PS-II)

The second pilot study was conducted in order to inform the design and implementation of the worksheets for the current study, and to see if there was evidence of effectiveness of the worksheets in increasing grammatical accuracy from pre-test to post-test; the worksheets are described in Section 3.6.6.2. The participants in PS-II (n=16) were from the same institution, taught by the same teacher-researcher, and taking an intensive version of the same course as the students in PS-I; the content, assignments, and goals of the class were the same as for the non-intensive course. Because of the small number of participants, no groups were constituted; all students received the same treatment.

The pilot study allowed for the refinement of the worksheets. For example, it seemed the students struggled to differentiate between grammatical errors, which they had been asked to record, and other types of errors, such as formatting, content, and spelling, which were not the focus of the worksheets. This information led to the decision in the current study to number the targeted grammar errors identified for attention. Without a control group, claims about effectiveness cannot

be made, but statistically significant differences from pre-test to delayed post-test suggested that further research was appropriate.

## 3.6 Data collection for the current study

### 3.6.1 Context

The university was selected as the site of the study for several reasons. Firstly, because I had worked at this institution for eight years prior to beginning my doctoral degree, I was familiar with this particular population, their English L2 language needs, and their first language. When researchers are proficient in study participants' L1, they are able to incorporate and analyse L1 data provided by participants without relying on translations. For example, such researchers can conduct interviews in participants' L1 if needed, and respond in real time to interviewees' comments, which could contribute positively to the quality of the data collected.

The sampling frame can be said to be Korean first-year STEM majors in required English academic writing courses being taught by English L1 instructors at universities in Korea; it is acknowledged that the students at this particular institution tend to be high performers, perhaps limiting the generalisability of the study to similarly high-performing students in similar contexts. If it is the case that humanities students are more willing to speak in English, this increased level of output could result in greater learning, altering the outcome of this study (Swain, 1995). While it is possible that relevance of the study could be extended to similar contexts in other parts of East Asia or other parts of the world, this would need to be done with appropriate caution.

In order to limit the number of confounding variables, it was decided to use purposive sampling to constitute a homogeneous sample; the sample population was the entire first-year STEM cohort in required English writing classes at a high-ranking university in Korea, where entry into the institution is competitive.

#### *3.6.1.1 The university and the course*

The institution places students into compulsory first-year groups by intended field: humanities/ social science students take all of their compulsory courses together, and natural science students take all of their compulsory courses together. The two teachers who agreed to participate in the study were both teachers of English courses taken by natural science students, and so all of the participants in the study were intending to major in the natural sciences.

Of the English teachers at the institution where the current study took place, those who have experienced teaching both natural science and humanities groups (which are taught on separate campuses) anecdotally report that the natural science majors seem to have less willingness to communicate, lower interest in the English course, and greater dissatisfaction with their English classes and teachers than their social science / humanities peers following the same English curriculum at the same university. The Hong Kong university student participants in Peacock and Ho's (2003) study offered some possible reasons why students, particularly science, technology, engineering, and mathematics (STEM) students, might feel this way about English. A maths student in that study explained that they were 'not interested in English and just wanted to use simple words' when expressing themselves. A science major explained that the importance of other subjects overshadowed that of English. This seems not to be a unique situation; as discussed in Section 2.4.3.2, research suggests that STEM majors may have different tendencies and levels of interest in language learning compared to non-STEM majors (e.g. Wong and Nunan, 2011).

The course on which this study focuses is an academic writing course designed to help prepare Korean students for success in their academic careers, where approximately 30% of the courses they take for their majors will be delivered in English. Although this would appear to imply a learning-to-write context (Manchón, 2011), perhaps in part due to the low productive skill and lack of L1 and L2 writing experience (Sections 1.2.2 and 3.6.2) of the students, the course combines elements from both strands. The current study focuses on the *writing to learn (language)* aspect of the course. The university also wishes the class to help prepare students for English proficiency tests which will be required for future job applications. It is a for-credit course that is offered once each year, and a passing score is required for graduation. Students usually take the course in their first term at university. The students' GPA across all their first-year courses affects which majors they are allowed to select, a decision usually made just before entering their second year. For this reason, students tend to desire high marks regardless of their personal interest in or valuation of the subject, though this desire might not necessarily translate into effort, as will be seen in Chapter 6. Because of students' lack of L1 writing experience, the course begins with the most basic elements of writing, such as using appropriate margins and double-spacing, and progresses to the use of topic sentences and supporting ideas, all while working on students' grammatical accuracy. The expectation is that students will learn to produce structured paragraphs and essays, while at the same time improving their linguistic accuracy.

#### *3.6.1.2 The teachers in the current study*

The academic writing course in which this study took place has been a required course at this university since spring of 2005, and is taught by highly proficient English speakers, all of whom have at least a master's degree, though not necessarily in a related field. In general, each teacher teaches six sections of the class, each section having 20-22 students. Administrators decide which textbook will be used and provide basic guidelines and requirements for the syllabus, but teachers are generally free to decide how they wish to organise and teach the course. Perhaps being understanding of what the students had been through to get to their first year at university (Section 1.2.2), the institution tries to help students get good marks in their first year. Applying a cohort-referenced system to all of a given teacher's sections (i.e. each teacher's grading curve is independent of that of all other teachers), it implements a generous grading curve for the required English academic writing course focused on in the current study: up to 50% of students can receive an A or A+; up to 40% can receive a B or B+, and only 10% of students must take C+ or lower. Anecdotally, students expect that teachers will use all of the allotted A/A+ and B/B+ marks.

As explained above, I had previously taught at the institution where the current study took place. To recruit participants for the study, I approached former colleagues at the university who had five or more years of language teaching experience (eight of fifteen teachers), as research indicates that novice teachers might view CF differently than experienced teachers (Rahimi and Zhang, 2015); the choice of *five years* as the minimum experience is, admittedly, somewhat arbitrary, but it has been proposed that, although time is neither the only nor the most important factor, the number of years of experience to move out of the novice stage of teaching be require 2-5 years (Berliner, 1988). I explained the purpose and general design of my study, and asked if they would be willing to participate. Two teachers, English L1 males in their 30s, with eight and twelve years of teaching experience respectively, agreed to participate. Their nationalities were, respectively, English and American, and both had been working at the university where this study took place for approximately five years; one teacher's self-assessment of his Korean L2 proficiency was lower-intermediate; the other, high beginner. Therefore, both teachers have knowledge of the typological differences between students' L1 and L2, and could use this knowledge in their interactions with students. Both teachers had obtained permanent legal residency in Korea and, as stated verbally in informal conversations with me, thought of their teaching as a career rather than as a temporary job. Both had acquired *Teaching English as a Foreign Language* (TEFL) certification prior to beginning work at the university. In interviews conducted with the teachers after the intervention, both expressed enthusiasm for teaching grammar, referring to it as *systematic and interesting* (Teacher 1) and *fun, like a puzzle to be solved* (Teacher 2). That both teachers found value and enjoyment in

teaching grammar is important; if one of them resented the topic or greatly disliked it, then there is a chance that feeling could come through in their lessons, which could influence how the students in that teacher's class responded to the grammar prepared lessons. While the groups were counterbalanced to account for such variables, it is clearly ideal for there to be as many similarities as possible.

#### *3.6.1.3 The students in the current study*

First-year Korean STEM majors enter compulsory English academic writing courses at university with both a great need for English and great challenges in learning it. As explained in Section 1.2.2, their need for English is twofold: an immediate need exists because of the university courses which will be taken in English; and after university, English will be used for gateway purposes in further studies and in employment. The students are aware of these needs, but they face challenges in that they may be both mentally exhausted from their years-long preparation for the university entrance exam, (Section 1.2.2), and lacking in experience writing (Section 3.6.2).

The students were divided between twelve sections of the academic writing course; the two teachers each taught six sections. However, only five sections from each teacher were invited to participate in the study, as each teacher had a so-called *international class*, comprising primarily non-Korean students whose L1 and educational backgrounds differed from those of the Korean students, making them unrepresentative of the population. The two teachers had 195 students (80% male, reflecting the general trend in STEM subjects in Korean universities, WISET, 2015) in their combined ten participating sections, of whom 122 students (75% male) both qualified for the study and signed the consent form. Of the 73 students who did not participate, 56 consented but were excluded based on the criteria below; the most frequently occurring reason for exclusion was being under the age of 18 (n=25). The following criteria were used to determine eligibility for the study:

- first-year, first-term Korean L1 university students over the age of 18 (and thus legally able to give their consent to participate in the study; as many students were living in the dormitories, obtaining parental consent would have been problematic)
- who had not lived abroad nor attended an international school or specialist foreign language school
- and reported no knowledge of an Indo-European language other than English, and only elementary-level knowledge of either Japanese or Chinese
- if male, had not completed their compulsory military service, which is a two-year commitment (anecdotally, students mature greatly during this time of service and tend

approach their studies more diligently after completion of service compared to their peers who have not done their military service. Omitting these students helped to control the homogeneity of the sample)

Tables summarising the demographics and educational histories of the included participants by group are provided in Section 3.6.2. The participants had lower-intermediate level L2 proficiency according to a proficiency test I administered at the beginning of term, further explained in Section 3.6.3.2.2. Most reported commencing their English studies between the ages of 5 and 8 (40.2%) or between the ages of 9 and 12 (47.5%), with a few beginning earlier (2.5%) or later (9.8%). Most of the students reported no or very little experience writing extended-length essays (defined as 3 or more typed pages) even in their first language (77%). The test-focused education system in Korea, as discussed in Section 1.2.2, might explain the low levels of L1 writing experience seen in participants in the current study; East Asian participants in other research have similarly reported low levels of experience with L1 writing (Chandler, 2003). (The combination of many years of study and little experience using the language has been reported for other contexts in East Asia (Shintani and Ellis, 2015)). Most (93.4%) of the participants had either attended a science high school or had been on a science track in a standard high school. This information is important to understanding the students' educational backgrounds. Table 3-3 depicts the basic humanities and mathematics/science curricula of two high schools in Seoul, a mainstream high school (S.S.H.S., 2015) and a science high school (S.Sci.H.S., 2015). The differences are clear; the science high school requires many more science and mathematics courses, leaving less time for other studies such as foreign languages. This is not unique to Korea; research has indicated similar findings for modern languages and STEM students in UK universities (González-Becerra, in press).

**Table 3-3 A comparison of credits required of a mainstream high school and a science high school in Seoul**

| | Credits required by mainstream HS | Credits required by science HS |
|---|---|---|
| Social sciences/ humanities | 118 | 63 |
| Science/ mathematics | 38 | 117 |
| Total[1] | 156 | 180 |

[1]This is a summary of required humanities and science/mathematics courses at these two schools. Both schools have electives and other required elements in the curriculum.

### 3.6.2 Creation of intervention groups

The university where the study took place assigns students in groups of approximately twenty students each to *Learning Communities*, or *LCs*, for their first year, and these LCs take most of their first-year classes together. Each of the two teachers in the current study taught five sections of the

writing class, i.e. five LCs. Students from all ten LCs were represented in each of the three research conditions (Table 3-4) in order to reduce effects of variability that can naturally occur between classes.

Of the 195 total students who were invited, 122 both qualified and consented to participate, resulting in approximately forty participants in each of the three intervention groups. The remaining 73 students were, unbeknownst to them, put into a *non-participating group*. These students took part in class normally; they did all of the same work that the participants did (including the proficiency test and all pencil and paper tests, and writing assignments) and received the same feedback methods (worksheets or direct correction, but not conferences due to limitations on teachers' time). The only difference was that the data resulting from their writing samples and tests were not recorded. Participants, though aware of the study (Section 3.8.1), were not made aware of their intervention group assignment or otherwise informed that there were separate intervention groups.

To help control for demographic variables and to reduce variability resulting from teacher and LC, the intervention groups were stratified by LC, gender, educational history and English proficiency to create similarity across the three groups (Cohen, Manion and Morrison, 2011:154). This stratification was accomplished in four steps: 1) creating a spreadsheet of participants' relevant background information; 2) assigning students to groups without consideration of background variables; 3) checking the statistics for the relevant background features, and then 4) manually making adjustments, spreading out the individuals and their characteristics to create balanced groups. For example, if there were five females in one group and three in another, then one of the females was selected from the group with five and moved to the group with three, leaving four in each group. The participant to be moved was selected through trial and error, e.g. determining which individual could be moved with the least disruption to the statistics of the other characteristics (e.g. gender, L2 proficiency, L1 writing experience).

Table 3-4 through Table 3-10 below provide descriptive statistics about the participants and demonstrate the comparability of the intervention groups. A chi-square goodness-of-fit test was used to confirm that the differences between intervention groups were not significant (reported below each table). When there are fewer than five cases in a category, the assumptions for the test were not met, so the test was not conducted. Nonetheless, it will be seen that they appear similar. For space, *Reflective worksheet group* is shortened to *Worksheet group*.

**Table 3-4 Distribution of participants (n=122) in intervention groups by Learning Community (LC)**

| | LC 1 | LC 2 | LC 3 | LC 4 | LC 5 | LC 6 | LC 7 | LC 8 | LC 9 | LC 10 | Tot |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Conference group | 4 | 4 | 4 | 4 | 2 | 5 | 3 | 5 | 5 | 4 | 40 |
| Worksheet group | 4 | 7 | 2 | 3 | 4 | 5 | 4 | 4 | 5 | 3 | 41 |
| Comparison group | 5 | 5 | 2 | 5 | 4 | 5 | 4 | 5 | 3 | 3 | 41 |

**Table 3-5 Gender by intervention group**

| | Females | Males | Total |
|---|---|---|---|
| | Number (%) | Number (%) | |
| Conference group | 10 (25) | 30 (75) | 40 |
| Worksheet group | 10 (24.4) | 31 (75.6) | 41 |
| Comparison group | 10 (24.4) | 31 (75.6) | 41 |
| Total | 30 | 92 | 122 |

*Males*: $X^2(2) = .022$, $p = .989$

**Table 3-6 Age of commencing English study by intervention group**

| | Ages 1-4 | Ages 5-8 | Ages 9-12 | Ages 13+ | Total |
|---|---|---|---|---|---|
| | Number (%) | Number (%) | Number (%) | Number (%) | |
| Conference group | 1 (2.5) | 13 (32.5) | 20 (50) | 6 (15) | 40 |
| Worksheet group | 1 (2.4) | 21 (51.2) | 17 (41.5) | 2 (4.9) | 41 |
| Comparison group | 1 (2.4) | 15 (36.6) | 21 (51.2) | 4 (9.8) | 41 |
| Total | 3 | 49 | 58 | 12 | 122 |

*Ages 5-8*: $X^2(2) = 2.122$, $p = .346$; *Ages 9-12*: $X^2(2)=.448$, p=.799.

**Table 3-7 Knowledge of a foreign language other than English by intervention group**

| | No L3 | Some L3[a] | Total |
|---|---|---|---|
| | Number (%) | Number (%) | |
| Conference group | 34 (85) | 6 (15) | 40 |
| Worksheet group | 34 (82.9) | 7 (17.1) | 41 |
| Comparison group | 37 (90.2) | 4 (9.8) | 41 |
| Total | 105 | 17 | 122 |

[a]Some L3 = Elementary level proficiency or lower in either Chinese or Japanese ; *No L3:* $X^2(2)=.171$, p=.918.

**Table 3-8 Participants attending science high schools (or being on a science track in standard high schools) and those attending a standard high school by intervention group**

| | Science | General studies | Total |
|---|---|---|---|
| | Number (%) | Number (%) | |
| Conference group | 38 (95) | 2 (5) | 40 |
| Worksheet group | 39 (95.1) | 2 (4.9) | 41 |
| Comparison group | 37 (90.2) | 4 (9.8) | 41 |
| Total | 114 | 8 | 122 |

*Science:* $X^2(2)=.053$, p=.74

**Table 3-9 Number of extended-length[b] Korean L1 essays students report having written during high school career by intervention group**

| | None | 1-3 | 4-6 | 7+ | Total |
|---|---|---|---|---|---|
| | Number (%) | Number (%) | Number (%) | Number (%) | |
| Conference group | 20 (50) | 10 (25) | 6 (15) | 4 (10) | 40 |
| Worksheet group | 20 (48.8) | 12 (29.3) | 6 (14.6) | 3 (7.5) | 41 |
| Comparison group | 19 (46.3) | 13 (31.7) | 3 (7.3) | 6 (15) | 41 |
| Total | 59 | 35 | 15 | 13 | 122 |

[b]*Extended-length* was defined for the students as 3 or more typed pages;
*None*: $X^2(2)=.034$, p=.983; *1-3*: $X^2(2)=.400$, p=.819

**Table 3-10 Mean L2 proficiency scores (Oxford Online Placement Test) by intervention group**

| | Mean proficiency score (S.D.) | CEFR equivalency | Number of participants |
|---|---|---|---|
| Conference group | 66.7 (13.6) | low B2 | 40 |
| Worksheet group | 67.4 (15.5) | low B2 | 41 |
| Comparison group | 66.5 (16.6) | low B2 | 41 |
| | Mean: 66.7 | Overall: low B2 | Total: 122 |

*Proficiency*: $X^2(2)=.016$, p=.992

### 3.6.3 Instruments

Table 3-11 presents the ten instruments used to gather data for the study, with detailed information about each instrument. The purpose of each instrument was to gather information about the participants or the effects of the interventions, or to confirm fidelity of implementation. Following Derrick's (2016) recommendation for good practice when reporting instruments, information is provided, where appropriate, on inter-rater reliability and instrument reliability.

**Table 3-11 Instruments used in the study with an explanation of purpose and, where relevant, corresponding research question (RQ)**

| | Instrument | Purpose |
|---|---|---|
| **Instruments used to gather data for research questions** | | |
| 1 | Student writing produced at three time points during the interventions | To measure change in overall error ratios for RQs 1a, 3a, and 4a, and to measure change in accuracy of article use for RQs 1b, 3b and 4b |
| 2 | A researcher-designed self-efficacy measure | To provide a measure of self-efficacy for RQs 3a and 3b |
| 3 | Language aptitude test (Oxford University Classics Language Aptitude Test (1995)) | To provide a measure of language aptitude for RQs 3a and 3b |
| 4 | Reflective worksheets (for the Worksheet group) | To provide a measure of languaging for RQs 4a and 4b |
| 5 | Audio recorded semi-structured interviews with students and teachers | To gain a greater understanding of students' and teachers' experiences and perceptions for RQ5 |

**Instruments used to avoid confounding variables and confirm fidelity of implementation**

| | | |
|---|---|---|
| 6 | A researcher-designed online background questionnaire | To determine eligibility for the study and to allow the participants to be stratified into groups |
| 7 | The Oxford Online Placement Test (OOPT; an online test of English L2 proficiency; OUP, 2014) | To determine participant L2 proficiency, allowing the participants to be stratified into groups and facilitating comparability with other published studies |
| 8 | A researcher-designed pre-test of selected grammar points | To confirm that the grammar points selected for focus in the study did not form part of participants' current proficiency |
| 9 | Audio recordings of one-to-one conferences between students and teachers from the Conference group | To investigate fidelity of implementation |
| 10 | Institution-produced video recordings of the teachers (only) delivering grammar lessons | To investigate whether teachers' grammar lessons were implemented similarly; that is, to help avoid *teacher group* being a confounding variable |

#### *3.6.3.1 Instruments used to gather data for research questions*

##### 3.6.3.1.1 Student writing (Appendix A)

Samples of student writing served as the source for measurement of grammatical accuracy. As part of their normal class assignments, the students produced four paragraphs of new writing on related but different aspects of a topic (as in Van Beuningen, De Jong and Kuiken, 2012). These were marked according to a rubric that, for ethical reasons, omitted grammatical accuracy, as will be explained in Section 3.8.2. The first, third, and fourth paragraphs are considered as the pre-test, post-test, and delayed post-test, respectively. The second paragraph was not analysed; it provided an opportunity for a second iteration of feedback, as results in PS-I suggested that a single iteration of feedback might be insufficient for perceptible development of grammatical accuracy, as discussed in Section 3.5.1. The paragraphs were written in an academic, persuasive writing style following a prescribed format (i.e. topic sentence, supporting sentences, concluding sentence); the task was piloted in PS-I (Section 3.5.1). These four paragraphs were described to the students as *writing tests*, and were each written during one 75-minute class period. At the beginning of the first writing test, students were asked to choose from a list of four topics. The teachers reminded the students generally about what they had studied about paragraph organisation and then listed on the board the grammar points that had been studied in previous lessons, reminding the students to use these points in their writing. The participants in Van Beuningen et al.'s (2012) study were provided with vocabulary likely to be unknown, but as the participants in the current study had chosen their own topics, this was impractical. Thus, the students were given 10 minutes to think about the topic and consult their dictionaries and Internet via their smart phones, which the background survey had indicated every student possessed, and to write a maximum of 10 words on a provided note card. They could, for

example, make L1 or L2 notes of translations, figures, names, or dates, but they were asked not to write sentences. At the end of each writing test, the students submitted their notes along with completed writing, which the teacher then stapled together. This allowed me to ensure that students were using the notes appropriately.

The later writing sessions followed the same general procedure, but the students were instructed to write on a specific aspect of the topic they had chosen for the first paragraph rather than writing on a completely different topic. The students were informed of the aspect to focus on only at the beginning of each writing test, in order to prevent them from memorising scripted paragraphs before the test, as sometimes occurs in this context. Teachers reported that students tended to require the full class period to complete the paragraph. Following the discussion in the Literature review on the importance of revisions and the need for revisions to be requirements (rather than suggestions), all students then revised their work for resubmission after receiving feedback in the class following the writing session; revisions were not examined as part of the study, which focuses on the effect of feedback on new writing. The reason for this is that to take students' corrected work away from them and ask them to write revisions seemed pedagogically unsound (e.g. Shintani, Ellis and Suzuki, 2014), but to measure students' ability to copy their corrections from one page to another was not useful. Polio (2012:377) points out that '[it] is obvious that a writer can look at direct corrections and copy them onto a new piece of paper. Even if those corrections were indirect, the second piece of writing would be more grammatically accurate'.

As discussed in Section 3.6.2, a large proportion of students (77%) had little or no experience with academic writing even in their first language, and most (93%) had had limited language courses in secondary school, having either been in specialised science high schools or in science tracks in standard high schools. Therefore, while the OOPT indicates that the participants had an overall B2 level, the OOPT is not a productive test, and so it is reasonable to expect that the students might have struggled to produce academic L2 writing.

#### 3.6.3.1.2 Self-efficacy measure (Appendix B)

A measure of self-efficacy was included in the study so that this variable may be considered in the results, and to facilitate comparability with other research (e.g. Hsieh and Kang, 2010; Lee, Bong and Kim, 2014), where self-efficacy has been shown to influence learning, including language learning. Self-efficacy is described as 'people's beliefs in their capabilities to produce desired effects by their actions' (Bandura, 1997:vii). As intuition suggests, and research supports, mood can influence

participant reporting on a given instrument (Cohen, Towbes and Flocco, 1988). In the current study, however, self-efficacy (and other instrument) scores are not examined individually; given the number of participants in the current study (n=122), individual effects are likely minimised. Additionally, beliefs can develop over time (Stajkovic and Sommer, 2000). Stajkovic and Sommer (2000) explain that context is important for providing estimates of self-efficacy. As such, if students understood what they were being asked to measure, they might be able to provide more accurate estimates of their self-efficacy. Therefore, following previous research (Wood & Locke, 1987; Stajkovik & Sommer, 2000) students were given time – in this case, approximately half of a university term – to become familiar with the requirements of their academic writing class and activities, and then self-efficacy was measured at two time points: in Week 8, immediately prior to the intervention, and after the final writing sample had been completed.

Existing self-efficacy measures could be found for other language learning skills and for L1 or L2 writing, but none could be found that focused on the specific situation examined in this study: the ability to use learnt L2 grammar and use the feedback to improve new writing. Self-efficacy is not a general attribute; it is always reported for a specific situation, such as a task or skill (Bandura, 1984).Therefore, I designed a self-efficacy measure to fill this gap. To avoid problems as discussed in the literature review, effort was made to adhere as closely as possible to Bandura's (2006) guidance for creating self-efficacy measures. Below, Bandura's step-by-step guidance is outlined, along with an explanation of how the guidance was incorporated in the current measure.

- 'Efficacy items should accurately reflect the construct' (2006:308). That is, Bandura explains, it is important to phrase the items *in terms of perceived capability* (i.e. *can*) rather than, e.g. intention (i.e. *will*).

  *e.g. I can understand new grammar structures when they are explained to me in class.*

- Care should be taken when wording items to avoid testing other constructs. It is important that self-efficacy, a construct specific to a given situation, not be conflated with, e.g. self-confidence, which is a different construct. In the example above, the item focuses on students' ability to handle grammar explanations and feedback rather than, for example, on students' belief about their ability to succeed in the class or to master English grammar in general.
- Be comprehensive when breaking down the component aspects of what is being measured, and make sure each item relates directly to the domain. In order to demonstrate successful mastery of taught grammar an L2 writing class, one might need to: a) understand what is

being taught; b) use what has been taught; c) understand why the teacher corrected a given error; d) take the corrections and apply them to future work.

- Items should represent 'gradations of challenges or impediments to successful performance' (Bandura, 2006:311). As the participants in this study explained in the post-study interview, it is one thing to know about something, and quite another to use it. The steps above in point (3) aim to represent different challenges.
- Bandura (2006) explains that scales should begin at zero (reasoning that negative self-efficacy is not logical), and that the scales should stop at 100, and that the scale should be sliding rather than discrete, as discrete scales can reduce sensitivity. For the current study, participants were free to locate their self-efficacy anywhere on the scale from zero to one hundred.
- In order to help participants understand that the items are about ability rather than, for example, desire or intent, clear instructions and an example should be given. Thus, the instructions for this scale were provided in both Korean and English; the steps for arriving at an answer were explained, and an example was given.
- Bandura (2006) encourages researchers not to use names on self-efficacy scales, but due to the need to track the three groups, as well as the non-participants, who were taking all of the same tests as their peers but without having their scores recorded, this was deemed impractical. Thus, it is acknowledged that the fact that participants' names were on their papers might have influenced how they answered the questions on this measure (and, arguably, on all measures requiring students to identify themselves).
- For reliability, Bandura's advice is to obtain a measure of internal consistency using Cronbach's alpha. For the current study, the internal consistency of the self-efficacy test resulted in a Cronbach's alpha score of .824, which is considered good, particularly for a scale with only five items (Cortina, 1993), and which compares favourably with those reported in similar studies (e.g. Hsieh and Kang, 2010).
- Following Bandura's (2006) advice to pilot self-efficacy scales, the instrument was piloted at the same time as the grammar correction test, and with the same pilot-participants. Specifically, Bandura warns about problems with ambiguity and with large numbers of participants indicating the end points of the scale (that is, zero or 100). The piloting of the instrument demonstrated that students understood what to do, and thus, that the items seemed not to be ambiguous. The respondents exhibited a range of scores, and pilot participants seemed to understand what to do, so no changes were made to the instrument.

The self-efficacy instrument was administered by the teachers, who reported that no difficulties were noted during the administration, and that participants did not appear annoyed when asked to complete it again at the end of the term.

#### 3.6.3.1.3 Language aptitude test (Appendix C)

In order to further investigate in which situations language aptitude contributes to acquisition, and to increase the comparability of this study with others in the field where language aptitude is measured (as below), it was decided to include a measure of language aptitude in the current study. As discussed in the literature review, language aptitude can be described as the 'ability to reflect on language and extract abstract information' (Ellis, 2004:251), an ability which tends to be relatively stable (Skehan, 1998:187f). As this study focuses on grammatical accuracy, and following other research on similar topics (DeKeyser, 1993; Sheen, 2007; Kormos, 2012; Shintani and Ellis, 2015), it was decided, as discussed in the literature review, to operationalise language aptitude as *language analytic ability*.

There are several existing language aptitude tests which measure language analytic ability, such as: a) MLAT (Carroll and Sapon, 1959); b) the LLAMA Swansea Language Aptitude Tests (Meara, 2005); c) the PLAB (Pimsleur, 2014); d) a test developed by Schmitt, Dörnyei, Adolphs, and Durow (2004) specifically for their study; and e) the Oxford Language Aptitude Test (Langslow, 1995). The language aptitude portion of the sample MLAT on the University of Oxford's web page contains English vocabulary that is likely to be unknown, so it was rejected to prevent effects of vocabulary knowledge as a possible confounding variable. For the context of the study, an online test, such as Meara's LLAMA, could not be considered due to the lack of classroom computers; the room where the proficiency test had been administered was in high demand, and it was not possible to reserve the room for a second session. In addition, the LLAMA website warns that the test is 'very experimental, and should be handled with extreme caution' (Meara, 2005).

The PLAB is geared toward students in secondary school who are about to begin their study of a foreign language; additionally, the cost ($27 per test) was beyond the budget of this study. Schmitt et al.'s test is designed similarly to the MLAT example, but uses simpler English vocabulary; it seemed to be ideal for learners in this study and had been used in similar studies (Schmitt et al., 2004; Sheen, 2007). The test provides a gloss containing some artificial words and their English translations, such as:

- kau dog
- meu cat
- kau meud bo The dog is chasing the cat
- kau meud bi The dog was chasing the cat

Then the test gives the students an English sentence and four possible translations in the artificial language, from which students are asked to choose the correct answer.

- The dog is watching the cat.
  a) kau meud so
  b) kau meud si
  c) meu kaud so
  d) meu kaud si

Unfortunately, upon contacting the authors of the test, it was discovered that the answer sheet was not available. Attempts to contact other researchers who had used the test were also unsuccessful.

Like the above tests, the Oxford Language Aptitude Test also measures grammatical sensitivity. It is a paper-based test, designed similarly to Schmitt et al.'s (2004) test; it does not confound the test results by inadvertently testing learners' L2 vocabulary knowledge. Like Schmitt et al., it uses an artificial language to test students' ability to deduce rules. Additionally, it could be completed in the time available during a class. Due to its similarity with other tests, its affordability, and its practicality the Oxford Language Aptitude Test was chosen for this study. I contacted the author of the test in order to confirm the answers of the test to ensure reliability in marking (Langslow, 2014).

Below are some examples of the type of information given and questions asked on the test. A full copy of the test can be found in Appendix C. The test begins by providing a gloss containing some artificial words and their English translations, such as:

- hi-tiacumya-? 'Is a cat listening carefully?'
- hi-tisno-sist? 'Is the little girl listening sleepily?'

Students are then asked to analyse the information (provided only in part here) in order to identify various nouns, verbs, adverbs, and articles. For example, a student would be expected to infer from the full provision of information, that *mya* is the artificial language word for *cat*, and *acu* is the artificial language word for *carefully*. In this way, the students create their own gloss rather than have one provided for them, as was the case for Schmitt et al.'s (2004) test.

After the students are familiar with the nature of the activity, it becomes more challenging. The students are given four sentences in the artificial language, along with their English translations, and

then, by analysing those sentences, are asked to translate a fifth sentence from the artificial language into English:

- cunmati kid. 'A child is coming.'
- go-ti kid't. 'The child is going.'
- mu- kid'n go-pi. 'A cow and a child are going.'
- cunmapi ben mu-'t'n, la-pi'n. 'A boy and the cow are coming and singing.'

Translate into English:

- kid't mu-'n cunmapi, sno-'n go-pi.

  (*Answer: The child and a cow are coming, and sleepily (they) are going.*)

Students are also asked to translate from English into the artificial language:

- A cow is coming. (*Answer: Mu cunmati.)*

The test was piloted with the same group of 20 students as the grammar test and the self-efficacy test, but during another class, so as not to cause fatigue. I was initially worried that the test would be overly difficult and might frustrate the participants, so the teacher of the pilot group was asked to keep an eye on the students' affect and stop the test if they began looking stressed. Quite the opposite, the teacher reported that the participants enjoyed the test, seeming to view it as a puzzle. Thus, the decision was made to use this test in the current study. Similarly to the experience in the pilot study, the teacher-participants in the current study reported that the students seemed particularly to enjoy the task.

Enjoyment, however, is not a sufficient criterion for a test. Given the earlier discussion in the Grammar pre-test section on the problems with reliability testing with low scores, the decision was made to test the reliability of only the first section of the test, i.e. the first seven items (where the students are creating the gloss from provided sentences), as the test is designed to become progressively more difficult. As the answers were dichotomous, a Kuder-Richardson-20 test was performed, resulting in KR-20 value of .496, indicating low reliability with even the first few items; it can be assumed that whole-test reliability would produce lower reliability results; Erlam (2005) reported similar problems in her study, with low scores (MLAT, in this case) presumably resulting in the low reliability seen.

#### 3.6.3.1.4 Reflective worksheets (Appendix D)

Researcher-developed worksheets, conceptualised as explicit output-prompting CF, facilitated through self-scaffolding, were devised to guide participants in the second intervention group

(Reflective worksheet group) through the process of languaging, working through their errors independently (Section 2.4.4.3.2). Santos, López-Serrano, and Manchón (2010) describe a similar instrument they refer to as a *noticing table*. Participants in their study noted on the table a) the error and its correction; b) the error classification (e.g. grammar, lexis); and c) their acceptance/ rejection of the correction (and if the latter, why they had rejected it). The worksheets in the current study were similar in structure, but there was no need to classify the errors, as only grammatical errors were tagged for participant attention for the worksheets (Section 3.6.6.2), and no need for noting acceptance or rejection. Due to the social structure (Section 1.2.1), teachers' authority tends to be accepted, and it was thought that students might hesitate to reject corrections; analyses of the conference transcripts in Chapter 5 will demonstrate this to have been a correct assumption.

The primary purpose of the worksheets was to elicit metalinguistic reflections, engaging participants with the explicit knowledge of the targeted grammar forms. Participants were invited to use either Korean L1 or English L2 when completing the worksheets, following Lapkin, Swain, and Knouzi's (2008) finding that requiring learners to produce metalinguistic reflections in L2 independently can be difficult even at the intermediate level. The quality of reflection on the worksheets was measured, following the observation by Qi and Lapkin (2001) and Sachs and Polio (2007) that higher-quality noticing (that is, noticing with evidence of understanding) resulted in more corrections. Methods of analysis are described in Section 3.7.1.2. The reflective worksheets were piloted in PS-II (Section 3.5.2).

#### 3.6.3.1.5 Audio-recordings of interviews (Appendix E)

Approximately 10% (n=13) of student-participants were recruited for post-study interviews, which were included to permit consideration of 'the relevance of societal context' on development (Ahmed, 1994:157) by inquiring about participants' own experiences during the study. Interviews give researchers insight into insider (i.e. participants') views of instructional interventions, and allow the researcher to 'elicit additional data if initial answers are vague, incomplete, off-topic, or not specific enough' (Mackey and Gass, 2016: 225). Semi-structured interviews in particular give researchers 'freedom to digress and probe for more information (Mackey and Gass, 2016:225). Thus, recorded semi-structured interviews were selected to allow insight into the students' and teachers' perceptions of the tasks and outcomes, using, when appropriate, student writing, background surveys, and self-efficacy tests as stimuli (ibid.). Studies applying mixed-methods approaches, and which include interviews, tend not to invite views from those in the comparison or control group. It seems that the views of the group not receiving treatment are as valuable as those who are,

particularly if the participants were in a standard-treatment comparison group, as the ones in the current study (cf. Section 3.6.6.3). Thus, members from all three groups in the study, rather than just the intervention groups, were invited to participate in the interviews. Characteristics of the interviewees will be provided in Section 6.2.

Some researchers might emphasise the importance of obtaining objective truth in an interview through, for example, minimising researcher influence and bias, something Alvesson (2003:15) describes as 'imitating quantitative ideals'. Illustrating this, instructions for interviews often emphasise to readers the importance of eliminating bias, obtaining objectivity and achieving the most reliable and valid data (e.g. Cohen, Manion and Morrison, 2011; Robson and McCartan, 2016). Even when researchers concede the need for some flexibility in interviewing, they sometimes nonetheless simultaneously maintain a demand for sameness: Robson and McCartan (2016:291) call for some use of verbatim scripts in semi-structured interviews. However, researchers are increasingly calling for a more responsive, reflective qualitative inquiry (Koro-Ljungberg, 2004; Koro-Ljungberg et al., 2009; Qu and Dumay, 2011), and this is more responsive, reflective inquiry resonates with the study's pragmatic philosophical underpinning (*cf.* Section 3.3). Roulston (2011) argues that some traditional 'problems' can benefit interviews, such as using closed questions (traditionally shunned) as a resource for the purpose of eliciting specific information or confirming researcher understanding; and offering participants possible answers to questions (traditionally shunned) as a way of helping participants formulate their responses. An interview excerpt offered by Roulston (ibid.) shows the interviewee choosing among and elaborating on some of the offered answers while rejecting others. Brinkmann and Kvale (2015) embrace the researcher as the bearer of the craft of interviewing, arguing that the 'skills, sensitivity, and knowledge [of] the researcher [are] essential for the quality of the knowledge produced' (p73).

Even when flexibility is applied to interviews, validity is a necessary consideration. Oppenheim (1992: 96f) lists common threats to validity in interviewing; this list is commonly cited as guidelines for researchers (e.g. Cohen, Manion and Morrison, 2011:205). Below are Oppenheim's warnings, along with a brief discussion of how each applies to the current study.

- **Biased sampling**: Potential interviewees were firstly selected among those who had signalled their willingness to the interviewed on the consent form. Amongst these, effort was made to select those who had elaborated beyond the minimum when filling out the background survey at the beginning of the term, as many of the randomly-selected PS-I interviewees seemed shy and reticent. Thus, it was rationalised that students who had already demonstrated willingness to

share their thoughts might be more cooperative in interviews than students who had demonstrated that they were less willing. While there is a chance this might have skewed the sample to students who felt more strongly about English, either negatively or positively, it was important for the study to have interviewees who were willing to communicate beyond minimal answers.

Next, amongst those identified as possibly being more communicative, a stratified sample across all three groups in the study was selected in order to ensure that the insights gathered were as representative as possible of the larger sample. Effort was made to create a balance for classroom teacher, intervention group, and gender. To recruit the selected interviewees, the teachers were given the list of names, and they invited the students to attend the interview by speaking to them individually at the end of class. The students received an invitation with information about the interview provided in English L2 and in Korean L1. This bilingual document provided information such as the location of the interview, the expected time it would take, and an offer of both a free beverage from the café where the interviews were being conducted and a small gift certificate as a token of appreciation for their time.

All of students invited for a short interview accepted the invitation and signed up for an appointment at a specific time, providing their mobile number. They were given a choice of three days during the last week of the writing course, which was the week before the final exams for their other courses. Appointments for each day were available from 10 a.m. until 5 p.m. to ensure that the students' various schedules could be comfortably accommodated. I sent text messages to the students on the morning of their interview, reminding them of the time and location, and all of the students attended their interview.

The interviews took place during the final week, after the delayed post-test, and were the only part of the study that took place outside of the students' normal class time. Neither individual participant performance on the writing tests nor that of intervention groups, was known at the time of the interviews, thus avoiding obtaining biased samples of participants who had performed particularly well, or interviewing high- or low-performers differently. I also interviewed both of the teachers; this took place in a quiet room near their shared office.

- **Failing to establish rapport:** As explained in Section 3.5.1, Korean students can be shy and reticent when faced with the prospect of speaking to a non-Korean. Koreans are not alone in this discomfort with outsiders; Rubin and Rubin (2012:181) found that their Thai participants were

similarly more comfortable with Thai interviewers than with non-Thai interviewers, even those who spoke the interviewees' L1, suggesting that it was not merely language or culture barriers causing the discomfort. For this reason, a Korean interviewer might have been desirable, but funding limitations prevented this possibility from being considered. To ease participant discomfort, I offered participants the choice of being interviewed in English L2 or Korean L1 (or a combination of the two languages. Likewise, as the pilot study participants had demonstrated discomfort being in a quiet room alone with me, a foreigner, the need of creating a more comfortable atmosphere was clear. Thus, the decision was made to hold the interviews in an on-campus café near the location of the students' English classes. Though researchers have traditionally held such interviews in quiet rooms, it had seemed that the pilot study participants were intimidated by being alone in a quiet space with the non-Korean teacher-researcher; anecdotally, discomfort is a common reaction when an individual Korean finds themselves alone with a non-Korean. Upon hearing the request for an interview space, the café manager offered a table in the corner and readily agreed to give me control of the volume of the music so that it would not be too loud for the interviews.

To further encourage and put at ease potentially reticent participants to volunteer answers, the interview applied the ideas from Roulston (2011), as above, using suggested answers where necessary in order to keep the atmosphere relaxed and the participant at ease. As will be demonstrated in Section 6.2, the participants in these interviews, like the ones highlighted in Roulston (2011), showed willingness to select, reject, and add to the suggested answers, demonstrating that though the interviewer aided participants in arriving at answers, their responses were valid representations of the perceptions they were willing to share.

- **Altering the wording of the questions, or changing the order of the questions between interviewees**: To address this problem while still allowing for flexibility, as discussed above, the questions for collecting data were *semi*-structured. The wording of the interview questions could not always be the same; because of individual preferences, differing levels of English L2 proficiency, and the need to make interviewees comfortable, some interviews were conducted all or mostly in Korean L1 (n=3), some all or mostly in English L2 (n=9), and one using both languages in approximately equal amounts. With eight years' experience interacting with similar students at this institution, for those who wished to use English L2, I was able to 'pitch the level of [English] to the interviewee's level of competence [without] overshooting [or] undershooting the target' (Rubin and Rubin, 2012:186). Participants exhibiting high levels of discomfort were put at ease with additional introductory-type questions that could be easily understood (e.g. *Do*

*you like English? Were you happy about this semester?*), while it was possible to move more directly into the content with participants who seemed more at ease.

- **Eliciting information in a biased way or ineffectively handling difficult interviews**: The pilot study interviewees (n=6) demonstrated much reticence; of seventy-five participant utterances, over half (n=43) consisted only of one or two words. As open questions had not been successful in the pilot study, it was decided to elicit information from participants in stages. This meant engaging participants through series of closed questions leading to more open ones, as discussed in Roulston (2011). For example, in one interview with a particularly reticent participant, the interviewer asked a yes/no question about their seemingly high performance on a writing test (*Were you happy with your writing?*) rather than a more open question, such as *How did you feel about your writing?* While this was arguably a leading question, the participant readily answered this yes/ no question (No, she was not happy), and this opened up space for me to inquire about the reasons for this feeling, to which the participant was then able to answer in greater detail. This case will be discussed in more detail in Results Section 6.2.
- **Failing to use support materials effectively:** I brought with me copies of the participants' writing samples and of their self-efficacy tests in order to prompt students to recall their work from earlier in the term and to give them a chance to offer concrete examples based on that work. This was mostly successful, as evidenced by participants referring to their work throughout the interviews to illustrate their points with examples.
- **Failing to code data resulting from the interviews consistently or failing to include relevant data**: It is acknowledged that researcher bias can never be fully removed from any study in any field. However, careful transcription, transparency of the analysis process, and inter-coding was used to diminish the overall magnitude of this threat to validity, as will be discussed in Section 3.7.2.

Given the above discussion, while it is unhelpful to view interviews as potential sources of unbiased truths, it is equally unhelpful to take what participants say in interviews as the full extent of their *truth*, even as they perceive it. The interview takes place in a specific context during a brief moment of the participant's life; a different moment could reveal different things. '[T]he fact that [interviews] are artful productions, that they will often be shaped by concerns about self-presentation or persuasion, does not mean that they cannot be accurate representations' (Hammersley, 2003:124). This suggests that while interviews should not be viewed as sources of impartial evidence, they can have value for what they are – participant perceptions that they are willing and able to share.

#### *3.6.3.2 Instruments used to avoid confounding variables and confirm fidelity*

##### 3.6.3.2.1 Background questionnaire (Appendix F)

A background questionnaire was needed in order to contextualise the learners and their learning history. This knowledge helps to situate the learners in the study, contributing to the understanding of any outcomes. The background questionnaire also allowed me to identify students who met the criteria for participation, and to facilitate the formation of stratified treatment and control groups for the study (e.g. gender, type of high school attended), as seen in Section 3.6.2. The questionnaire was piloted during PS-I and was found to provide valuable information without overwhelming the students with too many questions. The piloting of the instrument allowed me to clarify some of the questions. For example, the original questionnaire asked whether students had been abroad. Based on answers to that question, the revised version divided the question into two parts, first asking whether students had *travelled* abroad, and then asking whether students have *lived* abroad. Another addition was the request for information about the length of time abroad in the case of answering *Yes* to those questions.

##### 3.6.3.2.2 English L2 proficiency test (Appendix G)

Proficiency was measured in order to allow the groups to be stratified. There are a number of ways proficiency is commonly measured in language research. As standardised tests tend to be both time-consuming and expensive, researchers have found alternative ways to estimate participant proficiency. One common method is to take the minimum requirements for entry to the class or university as the level of the whole cohort; this method, of course, fails to allow for the representation of participants who performed better than minimum on the entry test. Another method is to use an in-house test, such as those used to place students into levelled classes at language schools. This approach to measurement, however, limits comparison between studies, as in-house test scores cannot be compared between studies due to their inherent variability. Another method of measuring proficiency is to ask participants to self-report their most recent standardised proficiency test score. The disadvantages of this are that participants might not correctly recall their score, or that they might be tempted to inflate their reported score when reporting it. Other methods include administering discontinued or practice versions of proficiency tests (e.g. TOEFL) or asking students to self-report scores. One disadvantage of using discontinued versions of standardised tests is that TOEFL tests change from year to year, and it is usually unclear how old the TOEFL test being used is. Another disadvantage is that it is often unclear how the tests were marked, and whether the researcher had designed an answer key or had been provided with one, all of which could reduce comparability between studies.

For researchers who can pay for a test, one option is the Accuplacer (CollegeBoard, 2015). Much research has been done to assess the reliability and validity of this test, as indicated by the meta-analysis by Mattern and Packman (2009). However, while Accuplacer might have been a well-researched option for the current study, the cost of this test made it a financially unfeasible option. A more affordable alternative to the Accuplacer is the Oxford Online Placement Test (OOPT). The OOPT is an online English L2 placement test from Oxford University Press (OUP, 2014) and has been used in research as a measure of English L2 proficiency (e.g. Winters and O'Brien, 2013; Li and Post, 2014). There are two sections: *Use of English* and *Listening*. The numbers of questions in each section varies somewhat because, like Accuplacer, it is an adaptive test; that is, incorrect answers are followed by easier questions, while correct answers are followed by more challenging questions (OUP, 2014). *Use of English* contains approximately 30 questions. The OOPT website explains that this section tests 'vocabulary, grammar, and the understanding of meaning in a conversation' (ibid.). The listening section, comprising approximately 15 questions, tests 'understanding of meaning in a conversation' (ibid.). An external research team, Brooke, Aden, Al-Kuwari, Christopher, Ibrahim, Johnson, and Souyah (2012), compared the OOPT to Accuplacer, and reported a moderate overall correlation (.69).

The OOPT is automatically and immediately marked, and results are reported in a set of three scores, one for each section and an overall score. Pollitt (n.d.) explains that the overall score is designed to correspond to CEFR levels, based on an unpublished study, conducted during the piloting phase of the OOPT, involving 30 schools in 11 countries. Researchers asked teachers in the study to complete *Can Do* statements, based on the CEFR *Can Do* statements, for their students, who had completed the OOPT. The researchers then used the *Can Do* statements to match OOPT scores to CEFR levels.

Pollitt (n.d.) explains that to create equivalency between the CEFR and other proficiency tests, the Council of Europe recommends asking teachers to gauge the difficulty of the items, i.e. asking at what level a student should be able to answer a given question correctly. Politt reports that the idea of asking teachers to gauge difficulty was rejected because it was deemed unreliable by the OOPT researchers: '[T]here is ample evidence in the research literature that even trained judges are not able to [gauge item difficulty] well' (Pollitt, n.d.:n.p.). No citations are provided, but a review of research reveals support for this claim. Shepard's (1995) report gives an example from the field of mathematics where judges assigned a mathematical question to the *basic ability* category on a national test, but later evaluations of the test revealed that only 37% of students whose performance was at the basic level correctly answered the question. On a smaller scale, Impara and

Plake (1998) investigated item-difficulty judgement with late-primary school science teachers (n=26; 5 male) from ten schools. Prior to the administration of a national test, teachers were given questions from the test and asked to assess for each one 1) how well borderline students, i.e. those currently barely passing their class, would perform, and 2) how well the class as a whole would perform. After the students had completed the assessment, researchers compared teacher predictions to actual performance. Teachers had made a total of 1,300 estimates; for the borderline group, only 22.6% were accurate; for the whole group, 41% of were accurate. Researchers also found that prediction accuracy was not correlated (statistical test unspecified) with teaching experience (-.08 for group estimates; -.23 for borderline group estimates). Recent research on ways to improve the accuracy of item difficulty estimations suggests that the challenge of item difficulty evaluation has not been resolved (e.g. Attali et al., 2014). The author of the current study cannot claim that benchmarking to CEFR levels is reliable, but the OOPT serves its purpose to control for relative proficiency among the students in the current study as well as making the study comparable to other studies which have used the OOPT.

Unfortunately, as the test is not very old, there does not seem to be much independent research on its reliability and validity. The OOPT website offers two articles, discussed below. As the intended audience of these articles seems to be practitioners, many research-relevant details about the participants and methods of data collection and analyses are missing. Pollitt (n.d.) discusses the test-retest reliability of the OOPT, explaining that other measures of reliability cannot be applied to the test due to its adaptive nature. Students from five schools (n=128) participated in Pollitt's study (participant details not given), taking the test twice (details not given) to examine test-retest reliability. Cohen and Holliday (1982:93) would consider the correlation coefficients reported for each section by Pollitt (Use of English .87; Listening .82) and for the overall test, .91, to be high. However, much is unknown about the study, so its results must be interpreted with care. Validity is examined by Purpura (n.d.), but he discusses the issue in only vague terms, describing the steps that have been taken to ensure the validity of the OOPT. Nonetheless, all proficiency tests contain flaws, and the OOPT was financially feasible, efficient, and possible to implement on the university campus during class time. It is more objective than self-report scores and is accessible to other researchers who wish to compare the results of this study to others, something that could be difficult with an in-house test.

The OOPT was administered to assess the proficiency of the participants in order to allow the stratification of the intervention groups and to allow comparison with other studies which also

measure proficiency. The listening portion of the test can be administered in American English, British English, or a combination of both; some participants in PS-I had commented that they had found the British English accent distracting and more difficult, and no one commented that they preferred the British English version. Thus, the decision was made to use American English for the current test. The results of the background survey administered to the current participants in the current study confirmed that this was the right choice for the current participants; 79% of participants reported that North American accents were more familiar to them. Certificates of completion containing the test results were presented to students at the end of the study. The participants' overall OOPT score, rather than section scores, was used to gauge proficiency in order to ensure the comparability of the participants in the current study with participants in other studies, for whom total OOPT scores have been reported (Winters and O'Brien, 2015; Li and Post, 2015).

One problem was found that could not be resolved for either the pilot study or the current one: for any text shown on the screen, unfortunately, all of the computers were equipped with software that gave Korean L1 translations of individual words as students moused over them, and it was not possible to switch this function off. Since all of the computers had this software, all participants were able to benefit from it equally, but it is acknowledged that these automatic translations present an unavoidable threat to test validity and reduce comparability with other studies.

The classroom teacher and I administered the OOPT together early in the term during the participants' normal class time in a computer room reserved in the university library, seating the students at every other computer to reduce distractions. The room was pleasant, comfortable, and quiet, and the classroom teacher and I stood in the back of the room so that the computer screens were visible and we could ensure that students were not using the Internet to look up information. The participants finished the test within 75 minutes and were allowed to leave as they finished.

#### 3.6.3.2.3 Grammar correction pre-test (Appendix H)

A short grammar error-correction test was administered to participants at the beginning of the term in order to establish that the participants were not proficient in the use of the target features (Section 3.6.3.2.3); such tests have been used in similar research (Liu and Gleason, 2002; Sheen, 2011). As the targeted grammatical features were selected specifically for this study, an existing test was not available. The test was researcher-designed in order to focus solely on the grammatical features proposed for instruction in this study; it consisted of a short researcher-produced text (103 words) that contained thirteen errors. The number of errors was not specified to the participants.

The instructions were given in L2 English, and the text was narrative in style on a non-academic topic. The participants were instructed to find and correct any errors (crossing out for deletion, using a caret for addition); an example of how to do this was provided at the top of the test paper. The vocabulary and sentence structure were purposely kept simple so as not to intimidate the learners in their first week at university, and to allow the learners to focus on grammar rather than on comprehension.

The test was piloted once at the start of the current study with a group of students (n=20) in the same cohort as those participating in the current study, but who were in a non-participating teacher's class. The results indicated that the instructions for the test were sufficiently clear, that the test could be completed in a reasonable amount of time, and that the grammar points selected for focus did indeed present difficulty for the students, and so no changes were deemed necessary. The test answers are binary (correct/ incorrect), so following previous authors (Muranoi, 2000; Lyster, 2004; Fulcher, 2010), the Kuder-Richardson-20 (KR-20) was later used to explore the reliability of the grammar correction test taken by this piloting group. It is unfortunate that the results revealed that the test had very low reliability, resulting in a KR-20 value of .044. A problem with measures of reliability, including the Kuder-Richardson, is that they use variance as part of the calculation of the reliability score (Kuder and Richardson, 1937). Thus, when many answers are wrong (or right), there is less variance, and low reliability may result (May et al., 2009). As the instrument was designed to test the hypothesis that the grammatical features were unknown to the students, low scores were expected.

The participating teachers administered the test to participants during the first week of the term at the normal class time and in their normal classroom. The students quietly completed the test without discussing it with their peers or using their mobile phones. The students took approximately ten minutes to complete the task, and the teachers reported that no difficulties were noted during the administration. I marked the grammar correction test, deducting one point for each incorrect answer. Deductions were not made for non-errors that students marked as errors because the focus of the study was on the specific taught grammatical structures. Incorrect attempts at corrections and overlooked corrections were counted as incorrect answers. The resulting score was number correct out of thirteen. Scores ranged between zero and 10. The difference in performance between the three groups was not significant, as demonstrated by a one-way ANOVA; $F(9, 105)=1.454$; $p>.175$.

**Table 3-12 Grammar pre-test scores by group (out of 13)**

| | Grammar pre-test | | |
|---|---|---|---|
| | Mean | SD | Range |
| Conference group n=36 | 3.2 | 1.9 | 7 |
| Reflective worksheet group n=33 | 3.5 | 2.0 | 7 |
| Comparison group n=37 | 3.5 | 2.3 | 9 |

#### 3.6.3.2.4 Conference audio recordings (Appendix I)

Conferences were designed to guide participants in the first intervention group (Conference group) through the process of languaging, working through their errors with the support of the tutor (Section 3.4.2.1). Teachers were to elicit metalinguistic reflections, engaging participants with the explicit knowledge of the targeted grammar forms. Audio recordings of the one-to-one conferences between the teachers and students were originally intended to verify the content of the conferences for the purposes of confirming fidelity to condition. The conferences were recorded using the classroom teacher's mobile phone, and the teachers sent the files to me for analysis; teachers were later asked to delete the audio files in order to protect student privacy. Once the conferences had been transcribed, it was seen that they contained little student input, and that the conferences had been carried out differently from the prescribed protocol; this is discussed in detail in Chapter 5. Thus, the decision was made to analyse the content of the audio recordings of the conferences in order to gain understanding into how they were carried out by teachers and how students participated in them. Methods of analysis are described in Section 3.7.3. Conferences were piloted in PS-I (Section 3.5.1).

#### 3.6.3.2.5 Video recordings of grammar lessons

The video recordings of the grammar lessons were viewed to confirm the similarity of how the grammar lessons were given in the two classes. The university records teachers' lectures (the camera is trained on the lectern) and uploads them to the university's student access web page for students to use when revising lessons. These are freely accessible to students registered for the class and to the teacher, but they cannot be downloaded. The teacher-participants agreed to make the videos of their grammar classes available for me to view, and I viewed only the videos containing grammar lectures in order to ensure that the two teachers were implementing similar teaching methods.

### 3.6.4 Targeted grammar features

Recall from the literature review that CF might be more useful for reinforcing learnt structures than for teaching new ones. Thus, all structures targeted in this study were either explicitly taught in the classroom, or were known to have been taught in school, as explained next. Eight grammar features were selected for feedback in the current study. Five of the features, discussed below, were in the students' textbook; thus, university administrators expected teachers to cover these points during the term. The three remaining features emerged from a needs analysis conducted on work from previous cohorts. These three features were: (a) subject/verb agreement, (b) use of plurals, and (c) a small range of tenses which are taught at the elementary proficiency level in Korean schools: simple present, present continuous, simple past, and future. Though numerous possible grammar features from the needs analysis could have been selected, these three were chosen for the study because students are generally assumed to have a command of these elementary-level features by the time they begin university due to English being taught as part of the national high school curriculum (Section 1.2.2). However, needs analyses with prior years' students in this context as well as current students' performance on the grammar error-correction test indicate that this is not the case. This lack of command over elementary-level features is a common finding in L2 research (e.g. Ishikawa, 2013). Because these features had been taught prior to university, no additional teaching of these three points was provided in the current study.

Prior to the interventions, the teachers taught one lesson on each of five grammar points (listed below) during the first half of the term. I selected these five points from the course textbook based on needs analyses from prior years teaching students in this context; these were confirmed by students' performance on the grammar error-correction test. Due to the students' history of English instruction, as explained in Section 1.2.2, it is likely that participants had been exposed to these features, but had perhaps not been required to produce them, a possibility supported by statements made by interviewees after the study (to be examined in the qualitative results chapter).

a. *at*, *on*, and *in* for prepositions of time
   Students were taught the basic rules: *at* for specific times; *on* for specific days/ dates; *in* for a stretch of time that is unspecific). They also studied some of the common exceptions to these rules, e.g. *at night*, and memorised a list of seven common words which do not usually pair with these prepositions: *this, every, last, next, yesterday, today, tomorrow*.
b. adverbs of frequency (e.g. *always, often, never*)
   Students were taught the basic rules: place adverbs of frequency before most verbs, after *to be*, and between two verbs (e.g. *have never been*)

c. *most (of), some (of), all (of)*, etc.
   Students learnt to use *most* directly with a noun, e.g. *most cities*, and *most of* with pronouns (e.g. my, their) and with the adjectives *this, these, those*, and with the definite article *the*.
d. use of comparatives and superlatives
   Students studied basic rules, adding *–er / -est* to single syllable words, and using *more / most* with multi-syllable words.
e. definite and indefinite articles (with a limited number of uses with learnable rules)
   Students studied articles of first and second mention (e.g. There is a cat. The cat is black.), for general vs specific (e.g. Could you open *a/ the* window?), and for specific instances where *the* is used, e.g. parts of a building and for things in nature (*the floor, the roof; the stars, the sea*). They also studied the zero article for 'things in general', e.g. *I like apples* vs. *I like the apples from this tree.*

### 3.6.5 Procedure

#### *3.6.5.1 Teaching*

Grammar instruction was not a part of the intervention, but as explained above, prior to the intervention, students were given lessons as usual, including grammar lessons on the five grammar points listed above, which were taught in weeks 3-7 of the fifteen-week term (Table 3-13 below); no grammar teaching took place after Week 7. The lessons could be classified as isolated form-focused instruction, following the discussion in Section 2.3.1. As the learners were not highly proficient, isolated form-focused instruction could arguably benefit the learners by limiting the amount of information they are required to process (Section 2.3.1.2). The teachers worked with me and with each other in order to ensure that the content and delivery of their respective lessons were as similar as possible. They developed a common syllabus and taught their lessons using the same PowerPoint file and teaching materials. One teacher's classes were held in the mornings, and the other in the afternoons. Teachers did not dedicate any lesson time to the three grammar points that had been covered in secondary school.

#### *3.6.5.2 Writing tasks*

All of the writing tasks, the first of which took place in Week 7, followed the same procedure. At the beginning of the first designated writing class, teachers informed students of the four topic choices that were current issues being debated in Korean society (legalisation of gay marriage, regulation of antibiotics, local pollution problems, or an issue regarding legal rights over a disputed island territory) and then students were given ten minutes to use their mobiles to find details they wished to use in

their writing, such as dates or translations (All students had indicated on the background survey that they used a smartphone, and the WIFI provided by the institution was fast, free, and reliable.) After this period, all notes, books, and mobile devices were put away, and teachers handed out blank, lined papers. As they began their writing, teachers listed the five grammar points which had been taught on the whiteboard, reminding students to use the grammar in their writing. Students were given the remainder of the class (approximately one hour) to write one paragraph on the topic they had chosen. Though I had initially requested that students be instructed to use three of the five grammar points in each writing session in order to ensure that the points were used in the participants' writing, the teachers refused this request, as they feared frustrating the students. (In a context where teachers' annual contracts are awarded based solely on student evaluations, this was a considerable concern.) Having selected their topic of choice during the first writing session, topics assigned at the beginning of the following writing sessions focused on specific aspects of the initially chosen topic; again, students were not informed of these prior to the start of each writing session.

I held daily meetings with the teachers in order to discuss any issues that had arisen in their classes; an issue that was brought up following the first writing session was that the students had appeared stressed about being asked to write without any notes, so I prepared small notecards for students to use in all future writing sessions. During the initial ten-minute research period, students used these cards to make note of facts and translations in L1 and/or L2, but they were not allowed to write full sentences; teachers monitored the room during this time to ensure that this instruction was understood and followed. The notecards were stapled to students' written work when it was submitted at the end of each writing session so that it could be confirmed that students were using the tool as instructed; a review of students' notecards confirmed that they had done as requested. The notecards seemed to have a calming effect on the students, and they voiced no further concerns.

#### *3.6.5.3 Feedback*

Results from PS-I had indicated that at least two iterations of a given feedback treatment are needed in order to see improvement; more than two iterations of feedback would have been difficult to organise due to the time constraints of the term. Therefore, for the current study, each group received two iterations of the assigned feedback treatment. Following Ferris' (2006) suggestion that all writing should be marked by one researcher for consistency, immediately following each writing session, I marked the writing of all 240 students. Non-participants were included because, in order to prevent knowledge of the intervention group from unwittingly influencing my marking, the writing of whole classes was marked all together; the writing samples were divided by intervention

group only after the marking was completed. The marking focused solely on the eight grammar points described in Section 3.6.4. I then passed the students' work to the classroom teachers, who added comments about the content of the writing on all of the papers with the same type and colour of pen (dark green) so that the comments would look similar to the students' eyes. Because PS-II results had indicated that students had difficulty limiting their corrections to *grammatical errors* when other types of corrections (e.g. spelling, content) were also provided, I numbered the grammatical errors on the papers of all three intervention groups. This provided two main advantages: a) the worksheet group could easily identify which errors they needed to focus on in their worksheets, and b) the teachers could easily see which points to focus on in the conferences.

In the classes following the first and second writing sessions, which were completed in consecutive weeks, week 7 and week 8, the students in the two experimental groups received CF according to their intervention group. After each of the first two writing sessions, each Conference group participant was to receive a four- to five-minute one-to-one conference with the teacher, initially prompting the student to identify the problem, but moving to direct feedback if necessary; Reflective worksheet group participants were to receive direct feedback and completed a reflective task using the worksheet (to be described in Section 3.6.6.2). The Comparison group received direct CF.

In the class following the revision and resubmission of the second writing task, the third writing task (immediate post-test) was completed; grammar feedback was not given on this until the end of the study. Approximately three weeks after the third writing task, the fourth writing task (delayed post-test) was completed. Students were informed of their mark on each paragraph after they had completed the intervention for that writing task, i.e. the end of that class period. This is because studies have suggested that students sometimes focus on their mark to the exclusion of attending to the feedback provided along with the mark (Butler, 1988; Klenowski, 1995; Taras, 2001; Vardi, 2009). At the end of the study, direct WCF on the third and fourth writing tasks, which had not received grammar feedback because of the delayed post-test, was provided for the students in all three conditions. An overview of the study is given in Table 3-13.

**Table 3-13 Summary of the study**

| | Session 1 | Session 2 |
|---|---|---|
| Weeks 1-2 | Introductory activities, consent, administration of proficiency, language analytic and proficiency tests, and grammar pre-test | |
| Weeks 3-6 | Grammar and English sentence/ paragraph lessons | |
| Week 7 | Grammar and English sentence/ paragraph lessons (cont'd) | 1st writing sample (pre-test) |
| Week 8 | **First CF intervention** by group and self-efficacy test | 2nd writing task (not included in analyses) |
| Week 9 | **Second CF intervention** by group | 3rd writing sample (immediate post-test) |
| Weeks 10-13 | Lessons on essay construction | |
| Week 14 | Lessons on essay construction (cont'd) | 4th writing sample (delayed post-test) |
| Week 15 | Self-efficacy test and post-study feedback on third and fourth writing samples, and interviews | |

When scheduling the study, the teachers' needs, the institutional schedule (e.g. avoiding tests during weeks when there was a holiday), and the need for sufficient time between the second iteration of feedback and the delayed post-test meant that there was little flexibility with scheduling. The timing of feedback delivery needed to be similar for all students; spreading it out over several classes would mean some students would receive their feedback further from the time of writing than others; the more time that passes between writing and feedback, the more trouble students may have effectively using it (Price et al., 2010).

A further institutional consideration was necessary. The curriculum set by the institution dictated that there should be a minimum of three marked paragraphs, and this study required four. The teachers were hesitant to give students more work, lest the students compare their workload to that of students in other classes and complain on end-of-term course evaluations. In order to achieve the goals of the study while respecting the teachers' valid concerns, the decision was made to tell students that the first paragraph was an opportunity for practice so that they would know what level of writing would be expected of them, and that the paragraph would be marked but the grade not recorded; all other writing tasks counted toward students' final marks. The teachers reported that the students seemed to value the opportunity and take the task seriously. Having outlined the study design, the tasks, and the intervention groups, the next sections will explain each intervention in turn.

### 3.6.6 The interventions

The theories underpinning the study were provided in Section 2.3.1. This section will demonstrate how the interventions link to the theories, and then offer a description of each intervention in the study. Section 2.3.1.3 introduced Gass' (2018) five stages learners go through in the process of learning new information. Bitchener and Storch (2016) have linked these stages to the provision and use of CF. They explain:

1. Apperceived input: Corrections on learners' written work prompts them to notice errors. Feedback for all participants in the current study was hypothesised to prompt noticing (Schmidt, 1990; 1995).
2. Comprehended input: The correction is understood; this may be a process of understanding rather than a comprehensive understanding. The CF treatments in the current study were hypothesised to support participants' comprehension through facilitation.
3. Intake: The learner compares the information from the correction with their prior knowledge. Gass' inclusion of this stage of the process is supported by information processing theory, which suggests that learners' ability to take in information can be influenced by their levels of knowledge of and experience with what is being learnt. In the current study, all participants were provided with grammar instruction on the targeted grammatical features to ensure the opportunity for declarative knowledge of these forms. The facilitation provided in the two treatment groups in the study was hypothesised to support learners' intake by prompting them to reflect upon and produce metalinguistic reflections about the errors. Returning to Ellis' (2006) discussion of grammar, discussed in Section 2.3.2.1, the purpose of eliciting metalinguistic reflections from participants in this study is to help them internalise the form by confirming that learners have an understanding of the grammatical forms. Within a cognitive framework, this could include '[bringing] learners' attention to L2 versions of their interlanguage utterances and [heightening] their awareness of their own interlanguage system' (Pica, 1994:154); or giving students 'declarative crutches' to rely upon while they are proceduralising their knowledge of the grammar features in question (DeKeyser, 1998:49). It was hypothesised that this depth of processing (Craik and Lockhart, 1972; McLaughlin, Rossman and McLeod, 1983) would develop participants' explicit knowledge about the targeted grammatical forms, and that this could result in improved accuracy in the use of these grammatical forms in subsequent writing tasks, either because the explicit knowledge promoted noticing and/or provided negative evidence; or because it helped establish declarative knowledge as a basis for proceduralisation, in a focused way that allowed for resources to be efficiently employed in processing the information. The use of feedback given to participants in the Comparison

group was not facilitated, and they were not prompted to produce metalinguistic reflections.

4. Integration: The new information is used to confirm (if the information matches existing knowledge), or disconfirm hypotheses. Disconfirmation in particular presents an opportunity for the learner to form a new hypothesis. Where no correction is given, existing knowledge might be strengthened. Alternatively, the information might be stored for later use, awaiting later (dis)confirmation, or it might be lost.
5. Output: The second part of Ellis' (2006) definition of grammar instruction is to help learners' internalise the form by processing it through comprehension or production (Section 2.3.2.1). Having encouraged comprehension in *Intake* (Step 3), this stage provides learners with opportunities for output, which include both the revisions learners make to their prior work and new writing they produce. It is the 'overt manifestation of whether or not the learner has begun the process of developing new explicit knowledge' (Gass, 2018:n.p.). Learners might be hypothesised to use their explicit knowledge to monitor their writing, rereading their work to find and correct errors prior to submitting the work (Sections 2.2 and 2.3.1.3).

The above five stages of the process represent a single 'episode' of CF, but the process is not intended to end at this point, as indicated by references to later learning throughout. Skill acquisition theory, as explained in the literature review, suggests that cognitive skills, such as writing and using grammar, can be developed through practice. All participants were given the opportunity to produce multiple iterations of writing and receive two iterations of feedback. The three interventions are summarised in Table 3-14, and then explained in turn.

**Table 3-14 Summary of the three interventions with explanations**

| **Group** | **Intervention** (provided twice each) | **Explanation** |
|---|---|---|
| Conferencing group | Facilitated with scaffolded metalinguistic reflections | One-to-one conferences |
| Reflective worksheet group | Facilitated with self-scaffolded metalinguistic reflections | Reflective worksheets |
| Comparison group | Direct WCF (unfacilitated) | Direct WCF |

#### *3.6.6.1 One-to-one Conference Group*

Conferences were conceptualised as *facilitating use of explicit output-prompting CF through expert-provided scaffolding* (Section 2.4). This intervention was designed so that the teachers would orally review targeted errors in student writing with the students. To ensure consistency, I provided direct written CF on photocopies of student work for the teacher to refer to during the conferences,

leaving students' original work clean, to be written on during their one-to-one conferences. The errors were numbered for the teacher's convenience and to mirror the worksheet condition (explained next). As a result of PS-I (Section 3.5.1), two decisions were made about the conferences. Firstly, in order to help students feel at ease, the conferences took place in the classroom with the other class members present (but busy working on their own assignments). Secondly, as students had shown reticence during the open, conversational-style conferences in PS-I, a more controlled protocol was developed for the conferences; this protocol and its development were explained in Section 3.4.2.1. The teacher was asked to call each student in the Conference group to the side of the room one at a time for a quiet four- to five-minute conference focusing on the grammatical aspects of the student's writing.

Teachers were asked to monitor students for signs of stress or frustration and react accordingly to temper negative emotions the student may have been experiencing. Even though this degree of flexibility in the protocol might introduce variation in the delivery of the conferences, it was deemed necessary to give teachers this flexibility for ethical reasons. After the conferences, the students were then to return to their desks and, like their classmates, revise their work. Issues relating to fidelity of implementation are addressed below in Section 3.6.6.4.

#### *3.6.6.2 Reflective Worksheet Group*

The Reflective worksheet group received worksheets, conceptualised as *facilitating use of explicit output-prompting CF through self-scaffolding* (Section 2.4), were designed to focus students' attention on their errors and prompt students to engage in languaging, producing metalinguistic reflections on errors marked on their work. In addition to an oral explanation about how to use the worksheets, an example was also provided. The worksheet was described in Section 3.6.3.1.4, and is provided in Appendix D.

Before giving the worksheets to the students, I prepared direct WCF (i.e., only error correction, not metalinguistic explanations) on the participants' errors and numbered them so that the errors corresponded to boxes on a worksheet. The errors were numbered because students in PS-II had had a difficult time differentiating grammar errors from other kinds of comments made by the teacher, such as content or formatting errors or requests for clarification. All worksheets included example explanations to show the students what was required of them, and this worksheet was returned, stapled to their paragraph. For each numbered error participants were asked to write a short explanation in Korean L1, English L2, or both. The students then revised their work. For ethical

reasons, if students asked specific questions about their work, the teachers answered those questions; implications of this will be discussed in the Results chapter.

The plan had been for the reflective group to receive, just before the next writing task, written feedback regarding the accuracy of their explanations on their self-explanation worksheets in order to mimic the scaffolding that was supplied to the Conference group. However, this plan proved to be unworkable because the students' L1 handwriting on the worksheet was often illegible to me. For the second iteration of the intervention, the students were asked to write more carefully, but the resulting worksheets were as illegible as those of the first iteration. The class did not meet between worksheet submission and the collection of the next writing sample (*cf.* Table 3-13), and it was not possible to decipher their writing and provide feedback in the short amount of time before the next writing. A Korean L1 user was needed to help work out what the students had written, a painstaking process that took several days even for the L1 speaker. The feedback was instead returned to the students after the completion of the study. This unavoidable deviation from the research design meant that while the participants in the Conference group received confirmation of the accuracy of their metalinguistic reflections, the Reflective worksheet group received no such support; this issue will be revisited in the discussion of the quantitative analyses.

#### *3.6.6.3 Comparison Group*

For strong intervention research design, Graham and Harris (2014:102) recommend that researchers '[compare] the targeted writing intervention to a *credible* control/comparison condition' (emphasis mine), though they concede this can decrease the likelihood of a positive outcome for the study. The decision was made to use a *standard-treatment control group* rather than a *no-treatment control group* (Vogt, Gardner and Haeffele, 2012). It would have been preferable from a research standpoint to have had a control group receiving no treatment at all, as in PS-I, as this would allow for the comparison of any effects of the treatment against a group receiving no treatment, demonstrating whether exposure and practice alone could result in increase in grammatical accuracy (Lyster, 2004). However, the teachers were understandably uncomfortable about withholding feedback from some students while providing it to others, as research generally supports the positive effect of feedback for learning. For ethical reasons, then, it was decided to compromise with a minimal provision of feedback to the control group; that is, to use a standard-treatment control group (henceforth *Comparison group*).

In the same way as for the reflective group, I prepared direct WCF on the errors of the work done by the Comparison group participants, but without numbering the errors. As in other research using direct CF (e.g. Parreño, 2015) corrections were given to the Comparison group by: a) writing the correct word(s) above an error; b) crossing out words which were causing the sentence to be incorrect; and c) inserting necessary words or letters in order to make the sentence correct. The Comparison group received direct WCF, and were asked to silently examine their feedback for a 4-5 minute period prior to beginning their revisions; this length of time was chosen to match the time the conference group spent examining their writing together with the teacher, which resulted in controlling the variable *time on task*, as the participants of all three groups spent about the same amount of time focusing on their errors. PS-I had suggested a four-minute conference was enough to promote understanding and uptake of feedback. In the current study, during this four- to five-minute window, the conferencing group students did not receive their papers for the purpose of helping to control for the variable of time on task. Based on experience with the students in PS-I, it was anticipated that some of these students might feel nervous when they did not receive their paragraphs; so to ease the students' concerns, the teacher explained before distributing the paragraphs that although there was not enough time to discuss each student's paper with them individually, that talking with some students helped the teachers understand students' grasp of concepts more clearly and would serve to shape the teacher's explanations, and that being selected for a one-to-one discussion was unrelated to the quality of the students' work. PS-I results suggested that this would suffice to calm the students, and the teachers reported that it did seem to accomplish this purpose.

After four to five minutes had passed, the teacher instructed those not receiving conferences (i.e. the Worksheet and Comparison groups) to work silently on their worksheets and rewrites while conferences were conducted with the conferencing students in turn, with the order being rotated so that the same student was not consistently last and consequently receiving the least amount of revision time. This was shown to be an unnecessary precaution, as teachers reported that all students were able to finish their revisions before the end of class. The Comparison group received these direct corrections of their errors without further elaboration or explanation, though to comply with ethical practices, the teachers answered any specific questions they asked. They revised their work for resubmission as the other groups had done.

#### *3.6.6.4 Fidelity of implementation (FoI)*

Researcher awareness of the importance of *fidelity of implementation*, also called *fidelity to condition, fidelity to protocol*, or *treatment integrity* has been increasing, and though the precise

definition varies from study to study, it is generally agreed that FoI considers the extent or degree to which a programme, intervention, or treatment is implemented as the developers of that programme, intervention, or treatment intended (Gresham et al., 2000; O'Donnell, 2008; Abry et al., 2013). A review of the literature suggests that *fidelity of implementation* seems to be the most commonly used term, so it is the one selected for this study. As was seen in the previous two sections, I maintained much control over two of the interventions, the worksheet intervention and the comparison intervention. Because the conference intervention was conducted not by me but by two classroom teachers, fidelity of implementation was of particular concern. Gresham, MacMillan, Beebe-Frankenberger, and Bocian (2000) identify four factors which have been cited in the literature as providing the basis for guidelines for researchers for strengthening FoI in their studies (e.g. Leech and Onwuegbuzie, 2007; Johnson et al., 2006). In order to avoid problems with FoI, these four factors were taken into consideration in the design of the current study as follows:

1. **Complexity** *(how complex the intervention is)*

   The conference procedure consisted of two steps, as described in Section 3.6.6.1: 1) Prompting the student to identify the error, and 2) Prompting the student to recall the relevant rule. Though taken individually the steps might be considered commonplace in teacher interaction with students, the study required that the steps be followed consistently and in order. Thus, to familiarise teachers with these steps, a document outlining the two-step procedure was sent to teachers in advance. On the day before the first conferences, I met with the teachers to orally review the conference protocol. Teachers were given the opportunity to ask questions, and they indicated understanding of the protocol.
2. **Materials and resources required** *(the extent to which these are readily accessible)*

   The teachers were given an outline of the conference protocol to keep with them. The teachers also had marked copies of each student's work to use as guides for providing feedback during the conferences, as described in Section 3.6.6.1.
3. **Credibility** *(the extent to which the intervention is consistent with teacher's beliefs)*

   In interviews, both teachers in this study expressed enthusiasm for teaching grammar, and both indicated that they regularly included feedback on grammatical accuracy when marking student writing. They both routinely held informal conferences with students to discuss students' work.
4. **Interventionists** *(the level of experience and motivation of the person implementing the intervention)*

   As described in Section 3.6.1.2, the teachers were highly experienced, trained ELT professionals. Their actions before the start of term suggest that they were motivated to

> implement the intervention: they had gone to great care to carry out their contributions to the study in the ways I had instructed; for example, they worked together to develop a single PowerPoint, which they both used in their classes, and they included the grammar points I asked them to cover in their lessons prior to the interventions themselves. They both redesigned their course syllabus and marking criteria around the needs of the current study.

There is an argument for me, as the researcher, to personally carry out these conferences, but the change from classroom teacher to researcher might have caused students to respond to the conferences differently than they normally would (Gravetter and Forzano, 2016:172). For example, students might have felt intimidated being in a one-to-one situation with an unknown foreigner (*cf.* Section 1.2.2), or, conversely, the novelty of the change from their normal classroom teacher might have prompted increased interest in the conferences. Alternatively, the change from their normal classroom teacher might have caused the students to view the conferences as separate from their class work, resulting in them discounting the importance of the conferences. They might, for example, make less effort to remember information provided in the conferences.

Another option might have been for me to observe the conferences, but concerns about affecting student response influenced my decision not to monitor the conferences in person, relying instead on audio recordings. My presence, as someone who was unknown to the students, might have caused the students to respond differently, either positively or negatively, to the conferences. For example, being watched by someone in the privacy of one-to-one conferences might have caused students anxiety, preventing them from focusing on the conference. On the other hand, being observed might have heightened their sense of how important the conferences were, causing students to pay closer attention. While being audio recorded can cause some people to become shy or otherwise respond differently, attention was not drawn to the fact that these one-to-one discussions with the teacher would be any different from any other discussions that students might have had with teachers earlier in the term.

As the above options were eliminated, it was decided that I would listen to a sample of the recordings of the conferences (two per teacher) immediately after the first intervention to confirm that they were being implemented correctly. When this was done, it seemed that the teachers were following the protocol. It was only later, when the written transcripts had been prepared and were being analysed, that a lack of FoI became obvious; the conference protocol had not been followed.

The evidence of this will be discussed in Chapter 5 . Given this discovery of problems with fidelity of implementation, the following post-hoc research questions were added:

6 How did the two teachers implement the conference intervention?

7 How did students participate in the conference intervention?

8 How did students participate in the worksheet intervention?

9 What, if any, comparisons can be drawn between the two interventions?

## 3.7 Data analysis

The following sections will explain the method of data analysis for the data in the current document. First, the procedure for analysing the essays is explained, followed by an explanation of how the worksheets were marked. Following this, the methods of analysing the interviews are given. Lastly, the methods for the post-hoc analyses of the conferences are explained.

**Table 3-15 Table relating research questions to instruments and plan of analysis**

| | Research question | | Data analysed | Analysis technique |
|---|---|---|---|---|
| **Differential effects of interventions on outcomes** | | | | |
| 1 | What is the effect of feedback delivery mode on: | | | |
| | a | the ratio of all targeted errors per hundred words? | Student writing | Mixed ANOVA with all targeted error ratios (DV) and group (IV) |
| | b | the ratio of correct article use to all obligatory article use? | | Mixed ANOVA with article ratios (DV) and group (IV) |
| 2 | What is the effect of feedback delivery mode on the development of self-efficacy over time? | | Self-efficacy scale scores | Mixed ANOVA with self-efficacy tests (DV) and group (IV) |

**Mediating variables hypothesised to influence outcomes**

| | | | | |
|---|---|---|---|---|
| 3 | What are the effects of language analytic ability and initial self-efficacy on: | | Student writing, language analytic test scores, self-efficacy scores (initial administration) | |
| | a | the ratio of all targeted errors per hundred words? | | Multiple regression with language analytic ability and self-efficacy (IV) and pre-test – post-test gain scores of all targeted features (DV) |
| | b | the ratio of correct article use to all obligatory article use? | | Multiple regression with language analytic ability and self-efficacy (IV) and pre-test – post-test article gain scores (DV) |
| 4 | For the worksheet group, what is the effect of the quality of reflection in the worksheets on: | | | |
| | a | the ratio of all targeted errors per hundred words? | Student writing and worksheet quality scores | Linear regression with worksheet scores (IV) and pre-test – post-test gain scores for all targeted grammar features (DV) |
| | b | the ratio of correct article use to all obligatory article use? | | Linear regression with worksheet scores (IV) and pre-test – post-test gain scores for article usage (DV) |

**Experiences**

| | | | | |
|---|---|---|---|---|
| 5 | Regarding the feedback delivery modes used in this study: | | | |
| | a | what are students' perceptions of their experiences? | Semi-structured interviews | Qualitative analysis of interview transcripts |
| | b | what are teachers' perceptions of their experiences? | | |

### 3.7.1 Analyses of data from the intervention

#### *3.7.1.1 Marking the essays*

The first step in analysing the writing samples was to decide how to count the words written, and to decide what was and was not an error. This was done by marking photocopies of a sample of the

essays to saturation (where no new marking rules were necessary). This iterative process resulted in the following guidelines:

- Repetitions of errors were counted separately (e.g. if the student used *sky* without the definite article three times in a writing sample, this was counted as three errors).
- Care was taken not to double-penalise students if one error led to another. For example, **Red apple is my favourite* counted as one error rather than two: *Red apples are my favourite*.
- All country names were counted as single words in order to create consistency between e.g. *the Republic of Korea*, *South Korea*, and *Korea*
- Names of people, together with any titles, were counted as single words because names, especially Korean names, can be written in a variety of ways (e.g. *Hong Gil Dong, Hong GilDong, Hong Gil-dong, Doctor/Dr Hong,* and *Hong*)
- Money and percentages were counted as one unit because of possible differences in expression, e.g. *$100* vs. *one hundred dollars* or *twenty (20) percent* vs. *20%*
- Hyphenated compounds were considered as two words because the pilot study showed that students tended to apply hyphens incorrectly (e.g. web-site, high-school). However, if a part of a compound could not be independent (e.g. non-governmental), it was counted as one word regardless of how it was written.

When this was completed, the actual writing samples were marked. Error ratios (numbers of errors per 100 words) were used to calculate the quantity of errors in participant writing; *cf*. Section 2.4.1. In addition to the concerns about t-units discussed there, another problem that motivated the use of error ratios was that the paragraphs in the pilot studies varied greatly in length. The participants generally did as instructed, writing 8-10 sentences for one paragraph, but the range of numbers of words in the first paragraph was 63 to 192 words – a wide range (mean 140 words across all four writing tasks). Considering the level of proficiency of student writing in the current study and the identified weaknesses of t-units and clauses, the decision was made in the current study to follow these researchers; for each writing sample produced by the participants, the total number of errors was counted and divided by the number of words the participant had written, and then multiplied by 100 to create an error ratio score per 100 words that would allow the participants to be compared against each other. A mixed ANOVA was run using these scores to investigate change over the course of the term. An inter-rater marked a paragraph from 10% of students (12 paragraphs), with a resulting intraclass correlation coefficient of .935.

#### *3.7.1.2 Marking the worksheets*

The worksheets prompted students to reflect on their errors, as described in Section 3.6.6.2. As was seen, each numbered grammatical error on students' work corresponded to a box on a worksheet. The errors were numbered because students in the pilot study had a difficult time differentiating grammar errors from other kinds of comments made by the teacher, such as content errors or requests for clarification. All worksheets included example explanations to show the students what was required of them, and each worksheet was returned, stapled to the source paragraph. For each numbered error students were asked to write a short explanation of the relevant rule in Korean L1, English L2, or a combination of the languages. Though the completed worksheets did not contribute towards students' final grades, but the teachers reported that students did not inquire about this. The worksheets were then coded for the level of languaging, hereafter referred to as *quality*, following previous research (e.g. Sachs and Polio, 2007). These worksheets were not returned to the participants until the end of the study (in contrast to what was originally planned; *cf.* Section 3.6.6.2).

Examples of coding of the quality of students' metalinguistic reflections were difficult to find in the literature; it seems that little has been done in this area. Some studies have avoided the problem by avoiding the consideration of quality altogether (Swain and Lapkin, 1998; Suzuki, 2012); others have chosen to maintain a very broad definition of quality, such as Sachs and Polio's (2007) two-tiered *Yes/No* classification system of quality: whether an error had been mentioned, and whether a change had been made, but it cannot account for levels of correctness, where attempts are close, but not quite sufficient. Roehr (2005) scored the quality of her students' metalinguistic reflections by differentiating between when participants used technical metalinguistic terminology (e.g. 'masculine accusative ending needed'), and when they described the issue without terminology (e.g. 'wrong ending'). A potential problem is that there seems to be a very fine practical difference between a student *describing* that a verb form is incorrect 'because they are still living there' or *explaining* that a verb form is 'to be used when an ongoing event/ action is being described'. Additionally, both of these represent alternative and effective ways for a student to demonstrate their awareness and understanding of the problem. Pragmatism, the philosophical root of this study, dictates that knowledge be conceptualised not as an objective reality, but as a social entity (cf. Section 3.3). Thus, if a student is better able to conceptualise a grammatical rule with a description, then pragmatism would not judge that socially-constructed knowledge to be inferior to a more 'objective' memorised explanation of the rule.

Kuikin and Vedder (2002) and Storch (2008) both used dichotomous coding schemes to gauge the quality of the languaging of their participants. Kuikin and Vedder, looking at oral interaction in small

groups, called their two categories *Simple* (i.e. mentioning a problem without discussing it) and *Elaborate*, which involves mentioning the problem, asking questions, stating the grammar rule, and making the change. Storch (2008) recorded her students in pairs engaging in languaging, and coded the talk as either *Elaborate*, defined as talk which involved deliberation, confirmation, or offering alternative answers, or *Limited*, meaning that the student offered a linguistic form as an answer with no ensuing discussion. However, with both of these systems, it seems that there is much room for possible engagement with feedback that falls between these two extremes. Kuikin and Vedder's classification breaks down into a) *No discussion* and b) *Much discussion with questions, stated rules, and application of the rule*. Storch's classifications have a somewhat less extensive gap between them, in that *Elaborate* means any *one* of deliberation, confirmation, or offering alternatives, and *Limited* means no discussion. Neither of these classification systems suitably accommodates situations in the current study where a participant provides either a partially correct self-explanation or no self-explanation at all. Because Storch's (2008) coding scheme has a smaller gap between two classifications than does the one used by Kuikin and Vedder, it was decided that Storch's coding classification scheme could be modified for the current study.

The first modification of the scheme was to add a *Null* classification to account for missing or inaccurate self-explanations, which seemed more appropriate than *Limited* for such situations. *Elaborate* was used to describe self-explanations which, keeping in line with the pragmatic foundation of the study, could include correct declarations of the rule, examples demonstrating the rule, or descriptions of the rule. *Limited* was used for self-explanations (the previously described declarations, examples, or descriptions) which were partially correct or where participants' understanding of the rule was unclear. Because worksheet quality was to be measured and quantitatively analysed, the second modification was to assign a numerical value to the classifications; thus, each *Elaborate* self-explanation on the worksheet was assigned a value of 2; each *Limited* self-explanation was assigned a value of 1, and each *Null* self-explanation was assigned a value of zero. The worksheet quality scores were confirmed with an inter-rater, resulting in an intraclass correlation coefficient of .926. Examples of the coding follow. Translations of Korean L1 self-explanations are provided in brackets.

a) Elaborate: 2 points

In the example below, the participant has correctly demonstrated their understanding that the definite article *the* is used for unique nouns.

| **Error** | **Correction** | **Metalinguistic reflection** | **Researcher's comment** |
|---|---|---|---|
| *Sky's color is* | *The sky's color* | *sky is only one* | *(Tick)* |

b) Limited: 1 point

In the example below, the student demonstrates understanding of using *an* with singular nouns, but seems to have failed to have conceptualised the *one of many* aspect of the indefinite article. Some singular nouns will use the indefinite article, so the self-explanation is close but incomplete.

| **Error** | **Correction** | **Metalinguistic reflection** | **Researcher's comment** |
|---|---|---|---|
| *if you have infectious disease* | *an infectious disease* | *Article 필수! Disease 는 단수임으로 an 사용. [Needs an article! Disease is singular, so it should have an]* | *There are many possible infectious diseases; here you mean one of those.* |

c) Null: 0 points

In the example below, the student writes that the word *second* should be accompanied by *the*, inadequately demonstrating their understanding of the definite article being used for unique nouns.

| **Error** | **Correction** | **Metalinguistic reflection** | **Researcher's comment** |
|---|---|---|---|
| *second* | *the second* | *The 필수 [Needs <u>the</u>]* | *There is only one second reason. It is specific, so use <u>the</u>* |

The following is an example of the worksheet score calculation process:

| | | | | |
|---|---|---|---|---|
| Elaborate self-explanations | @ 2 point each | 3 | = | 6 |
| Limited self-explanations | @ 1 point each | 3 | = | 3 |
| Null self-explanations | @ 0 points each | 2 | = | 0 |
| | Total self-explanation points awarded | | | 9 |
| | Maximum possible points | | | 16 (8 self-explanations) |
| | Final score | | | 9/16 = 56% |

### 3.7.2 Interview transcript analysis

Answering this final research question (RQ5), a content-oriented question designed for the purpose of triangulation and consideration of stakeholder perceptions, did not require a very fine-grained analysis of, for example, pause lengths and underlying meanings. Rather, the interviews sought to elicit more straightforward data – the perceptions that participants were able and willing to share. Thematic analysis is 'a method for identifying, analysing and reporting patterns (themes) within data

… which reports experiences, meanings and the reality of participants' (Braun and Clarke, 2006:79, 81). An advantage of thematic analysis is that it is independent of a given theoretical approach, and can fit well within different philosophical frameworks (Braun and Clarke, 2006). Section 3.3 provided a discussion of the philosophical and methodological orientation of the study; the flexibility offered by pragmatism's view of knowledge as a social entity and of scientists as agents striving to solve real-world problems (Talisse and Aikin, 2011) fits well with the flexibility also exhibited in thematic analysis, demonstrated in the number of choices available to researchers according the research questions they need to answer.

Braun and Clarke (2006) specify a number of decisions that researchers must make in the process of conducting thematic analysis, and they emphasise the importance of these decisions being made explicit, and with understanding of the implications of choosing one path over another. Table 3-16 summarises the key decisions that must be made; decisions made for the current study are shaded. A more detailed discussion about each of these choices will follow.

**Table 3-16 Choices researchers make in thematic analysis (Braun & Clarke 2006)**

| **Analyse dataset as a whole** | **Analyse a particular aspect of the dataset** |
|---|---|
| • Provides an overview of the themes from the dataset as a whole | • Provides a detailed and nuanced view of particular aspects of the data |
| **Develop themes inductively** | **Develop themes deductively (theoretically)** |
| • Some resemblance to grounded theory<br>• Themes tend not to be directly connected to the questions asked | • Themes may develop from questions asked |
| **Conduct a semantic analysis** | **Conduct a latent analysis** |
| • Accepts the surface-level meaning of what participants say | • Explores ideas and assumptions behind what participants say |

*Analyse the dataset as a whole vs a particular aspect*

The research question asks about participants' perceptions, so in order to gain understanding of these, themes from the dataset as a whole were included, rather than focusing only on specific parts. As Braun and Clarke (2006:82) explain, a 'theme captures something important about the data in relation to the research question, and represents some level of patterned response or meaning within the data set'. This allows the analysis to provide a sense of themes from the overall dataset, but this is not related to the frequency with which the themes appear. Braun and Clarke discourage researchers from focusing on frequencies of appearances of data relating to themes, as what makes

a theme is 'not necessarily dependent on quantifiable measures – but rather on whether it captures something important in relation to the overall research question' (p. 82).

*Inductive vs deductive coding of the data*

Braun and Clarke (2006) identify two ways data can be themed: a) *inductively*, meaning that the researcher examines the data without preconceived ideas about what might emerge from it, or b) *deductively*, meaning that the identification of themes is 'driven by the researcher's theoretical or analytic interest in the area' (ibid.:84); the authors also refer to this as *theoretical*, as it is driven by the researcher. I interviewed the participants in order to gain insight about specific experiences, and my review of the literature led me to focus on certain themes, some of which were clearly signposted in the interview questions; therefore, it can be said that the coding approach was a deductive one.

*Semantic vs latent analysis*

The decision to conduct a semantic analysis was inherent in the research question, which sought answers about the perceptions that participants were able and willing to share. The research question focuses on participant perceptions were of the intervention, and accepts participants' answers as representative of what they wish to share of their perceptions. As with any interview data, it is accepted that participants might have withheld some of their ideas, or framed what they did share in a way that portrays the ideas differently from their experience (Flick, 2014).

**Analysis procedure**

All of the audio-recorded interviews were transcribed, and a Korean L1 colleague confirmed the accuracy of the transcription of any Korean used in the transcriptions. Transcription was done using Microsoft Word while the events were fairly recent, within a maximum of approximately two months of the interviews. The transcripts 'just include the actual words spoken … [along with] nonverbal events that might influence the interpretation, such as when an interviewee laughs or shrugs' (Rubin and Rubin, 2012:191).

The data were then themed and organised into main ideas. The issue of reliability in studies with similar questions is not always addressed (I. Lee, 2008b; G. Lee and Schallert, 2008), but there are methods of enhancing reliability in cases like these, such as using triangulation of data and respondent validation (Lincoln and Guba, 1985; Beuving and de Vries, 2014). An example of these being used in similar research was found in Hyland (1998). Another technique found in research is

moderation (Woore, 2011). Thus, reliability of the analysis of the interview data in the current study was enhanced using these three methods. For triangulation, existing research, background survey data, and interviewee data were compared for inconsistencies. For respondent validation, one of the teachers from the study reviewed the analysis of the data (including student data), offering his 'evaluation of its authenticity to correct researcher bias' (Hyland, 1998:260). The teacher confirmed that the analysis was in line with his memory and experience. The third method was moderation, in which an applied linguist reviewed one (10%) of the transcripts and compared it to the themes drawn out by the researcher. The moderator had three comments: firstly, she noted that she agreed that all of those themes were prominent in the transcript. Secondly, she pointed out that the interviewee discussed his CF intervention without specifically referring to 'Reflective worksheets', though as the participants' attention was not drawn to the differences in treatments, this might be expected. Thirdly, she commented that there was not a theme to capture what interviewees said about non-English prior learning experiences. While this is true, and may have been interesting to explore, it did not have immediate relevance to the current study. Even with these checks for reliability, it is acknowledged that bias may exist, stemming both from the interviewees' decisions about what to share and from the researcher's interpretation of the data.

### 3.7.3 Post-hoc conference transcript analysis

A common analysis technique in qualitative research is discourse analysis (DA), which has been called 'a teeming and heterogeneous field of research' (Vaughan, 2012:280). As this description suggests, Vaughan (2012) explains that there are various approaches to DA; of relevance to the current study is the Birmingham model of DA (Vaughan, 2012), based on Sinclair and Coulthard's (1975) study of classroom interaction. This model of DA is concerned with speech acts; that is, language which 'is seen as performing a particular act', such as making a request (McCarthy, 1991:9). The current study is concerned with the speech acts of teachers and of students in the conferences. Research indicates that teachers make use of body language when communicating with learners about their writing (Thompson, 2009; Wang and Loewen, 2016); thus, it is acknowledged that since only audio data are collected, what is known of the interactions is incomplete.

Schreier (2012:6) outlines for researchers four steps which should be followed prior to the main analysis. Though she was referring to preparation for qualitative content analysis, these steps map readily, as Kuckartz (2014) suggests, to qualitative textual analysis more generally:

1. Building a coding frame
2. Dividing your material into units of coding
3. Trying out your coding frame
4. Evaluating and modifying your coding frame

#### *3.7.3.1 Step 1: Building a coding frame*

A review of literature was conducted in order to inform the building of the coding frame. One early analysis framework was *Flanders' Interaction Analysis Categories*, a ten-category system aiming to exhaustively code all classroom interaction (Flanders, 1970:34). Teachers' and pupils' speech formed the two main categories; their speech was then captured under the broad codes *Response* and *Initiation*. The *Teacher–Response* and *Teacher–Initiation* codes were both further divided into three sub-codes. *Pupil–Response* and *Pupil–Initiation* codes received no sub-codes, which severely limits the level of detail that can be captured by the coding structure (e.g. *How did the pupils respond?*). Another aspect of interaction, *Asks questions,* was classified neither in *Pupil* and nor in *Teacher* categories, leaving it unclear as to who was doing the asking. The category *Silence,* indicating lack of response, was conceptualised (curiously, given the IRF framework), as an independent code rather than a sub-code of *Response*. Nonetheless, *Silence* is a contribution to the classifications of the current study, as will be seen in Table 3-17.

Perhaps in an attempt to resolve the limitations of Flanders' system, above, Sinclair and Coulthard's (1975) hierarchical framework expands the number of categories to twenty-two, and is designed to capture even the most minute details of tutor-student classroom interactions, such as moves teachers use to signal the transition from one topic to another, for example *Well* or *Right*. This level of detail was not necessary for the research questions in the current study, which examine broader themes in the conference interactions. One result of Sinclair and Coulthard's (1975) work is the now well-known description of an *exchange* between pupils and teachers: initiation – response – feedback (IRF), with *initiation* being the domain of the teacher, and *response* that of the students. As it is, IRF fails to capture the complexity of interaction in the classroom, as exemplified by Mortimer's and Scott's (2003) IRFRF…., with an unlimited number of *R*s and *F*s following the I, which signal the continuation of the chain of further responses and feedback in tutor-student interaction. However, this model fails to accommodate initiation by students, or responses to student initiation by teachers.

The inadequacy of the IRF model in dealing with the complexity of exchanges can be seen in research attempting to apply this neat model to naturalistic and, therefore, messy interactions: Miao and Heining-Boynton (2011) applied the IRF model to classroom interactions between students and tutor in English-immersion kindergarten classes in China during whole-class activities. Like Mortimer

and Scott (2003), above, the researchers were forced to apply the model flexibly, identifying multiple instances of responses and feedback following a single initiation. Applying the IRF model in university settings, which might be expected to be more subdued and controlled, similarly resulted in insufficiently capturing the complexity of the interactions when Hardman (2016) applied the model to an undergraduate and a postgraduate class in the UK. Finding the three categories of IRF insufficient, Hardman specified sub-classifications within each of these; for example, whether the tutor asked open or closed questions, and whether students gave brief or elaborate responses. He also deconstructed the feedback move of the IRF model, specifying how the tutors were providing feedback – simply acknowledging the student's response, praising the response, or negating the response, for example. The model seems to suggest that *initiation* is a) used only by teachers, and b) to elicit one answer which can be categorised as correct or incorrect; thus, in this view *feedback* is different from *response*. Therefore, IRF cannot capture instances where the teacher is prompting the student to elaborate on their thinking; it also makes it difficult to capture instances where students initiate questions or comments. These inadequacies have, perhaps, led some researchers to seek alternatives.

Tsui (1985), drawing on the strengths of earlier frameworks, such as those developed by Flanders (1970) and Sinclair and Coulthard (1975), developed the *Seventeen-Category System* (SCS) for analysing classroom *exchanges*, which she defines as a series of *acts*; the seventeen categories in the framework are seventeen possible acts. A strength of the SCS is that it removes the classification of *feedback*, which can be considered a *response*. This omission of a potentially confusing category may result in a more consistent coding of data, as the researcher does not need to decide whether a teacher's utterance is merely a response to the student or is providing feedback to the student. The framework is divided into three main levels, which identify: a) the speaker (student/ teacher); b) they type of utterance initiation/ response; and c) seventeen possible acts. Some of the acts are then divided into further sub-codes.

It is interesting to investigate the degree to which Tsui's (1985) SCS framework fits within the pragmatic approach of the study and the SCT-informed interventions. The analytical framework needs to capture as fully as possible the teachers' efforts to facilitate students' understanding of the grammar rules they had studied; and if efforts to facilitate students' understanding were lacking, then the framework needs to be able to take account of areas of absence of facilitation as well. It is not enough to note whether facilitation occurred; in order to answer the research questions, in

cases where facilitation did occur, the analytical framework needs to demonstrate *how* teachers were scaffolding students' understanding, and *how* the students were participating in this process.

#### *3.7.3.2 Step 2: Dividing your material into units of coding*

Using Tsui's (1985) SCS as the starting point for the development of a coding framework for the current study, the first step of the analysis of the one-to-one conferences was to read through the transcripts several times to get a broad view of the data and get a feel for how well it could be accommodated by Tsui's categorisations.

Tsui's highest-level codes, *Teacher Talk* and *Pupil Talk*, are assigned according to who is speaking; in Tsui, as in the current study, the distinction between *Teacher* and *Pupil* is unambiguous. The next level is *Initiate* and *Response.* Though Tsui herself did not define these two codes, she did state that this analytical framework was based on the work of Flanders (1970) and Sinclair and Coulthard (1975); thus it seems reasonable to apply definitions provided by these authors. Sinclair and Coulthard (1975:21) explicitly state that *Initiation* comes from the teacher, and *Response* comes from the student. This unidirectional view of information transmission does not fit Tsui's framework, where both teachers and students are seen as capable of initiating. Nor is it in harmony with the socio-cultural perspective, where students and teachers are seen as working together to establish understanding. Though Flanders (1970:35) concedes that teachers will likely initiate more than students, his definitions of these terms allow for students to participate in the interaction by *initiating*. Flanders explains that *Initiation* means 'to make the first move, to lead, to begin, to introduce an idea or concept for the first time, to express one's own will'; and *Respond* means 'to take action after an initiation, to counter, to amplify or react to ideas which have already been expressed, to conform, or even to comply to the will expressed by others' (1970:35). These definitions fit well both within Tsui's framework and within the SCT perspective and are therefore adopted as working definitions in the current study. Tsui's categories can be found in Appendix J (Tsui, 1985:12*ff*.).

#### *3.7.3.3 Step 3: Trying out your coding frame*

Tsui (1985) specifies that she designed the framework with future research in mind and expected her system to be adapted; indeed, examples of modification of Tsui's framework can be found in the literature (Kim and Elder, 2005; 2008; Lo and Macaro, 2015). Therefore, the next step in the current analysis was to modify Tsui's framework by omitting categories that do not apply to this study and

adding categories that initial readings of the transcripts suggest to be necessary. This initial framework was tested on a small sample of the data.

#### *3.7.3.4 Step 4: Evaluating and modifying your coding frame*

Continued piloting of the coding scheme on a portion of the data allowed further refinement until the existing codes seemed to adequately capture all important aspects of the data, and no new codes needed to be added. After completing the four steps, the final application of the fully developed coding frame to the data was accomplished, using the tools provided by NVivo 11 for Windows to code the data. Table 3-17 provides the modified framework applied to the data for this study.

**Table 3-17 Modified coding scheme for the current study**

| | | Acts | Sub-categorisations | Descriptions of categorisations in the current study | Examples |
|---|---|---|---|---|---|
| Teacher | Initiate | 1. Elicit | A. Targeted Qs<br>a) Error Q<br>b) UnderstandingQ<br><br>B. More info(rmation) | A.<br>a) Prompts student to identify an error<br>b) Elicits student's understanding of rule<br><br>B. Teacher attempts to elicit information which would complete a student given answer | A.<br>a) *There's a problem on line 1. Can you find it?*<br>b) *Why* (is the answer we came up with correct)?<br><br>B. *Yes, that's the problem. Now how can you change it?* |
| | | 2. Inform | 2a) Inform-Error<br><br>2b)Inform-Explain | 2a) Teacher provides error correction or explanation to student<br>2b) Teacher provides error explanation to student | 2a) The problem is here, *Most of* – you don't need *of*.<br><br>2b) Use *Most* directly with nouns. Use *Most of* with another word, like a determiner or a pronoun. |
| | Response | 3. Evaluate | 3a) Accept<br>3b) Reject | 3a. Accepts pupil response<br>3b. Rejects pupil response in word or tone | 3a. *Yes, good*.<br>3b. *No, that's not it*. |
| | | 4. Reformulate | | 4. 'expand, develop, or provide additional information' to pupil's response (p15) | 4. Student: 'Because noun'.<br>Teacher: Yes, it's because *house* is a noun'. |
| | | 5. Redirect | | 5. Redirect student who has identified a non-targeted error | 5. *Yes, that's a spelling error, but let's focus here at the end of the line on a grammar problem*. |
| | | 6. Clue | 6a Indirect | 6a. When error has been identified but not corrected; guides student's thinking by giving prototypical examples | 6a. *I like* ___ (pretends to sip coffee to elicit null article with general noun *coffee*) |
| | | | 6b Direct | 6b. When error has been identified but not corrected; guides student's thinking by directly referring to the problem, either by pointing out relevant parts of the sentence or by referring to the conceptual framework used in teaching the relevant grammar point | 6b. *If this is your noun, then what do you need to put here?* (talking directly about the sentence) or *Is it specific of general?* (referring to the conceptual framework used in the lesson) |
| | | | 6c Narrowing focus | 6c. When student is unable to identify error; offers a clue about the type of grammar or more specific location | 6c. *Look at the end of the sentence* or *You're looking for an article problem*. |

(Shaded areas indicate *Initiate* categories; non-shaded areas indicate *Response.* Student codes on next page)

| | | | | | |
|---|---|---|---|---|---|
| Student | Response | 7. Reply | 7a) Restricted<br><br>7b) Silent<br>7c) Expanded<br><br>7d) Indistinct<br><br>7e) Uncertainty<br>7f) Silence | 7a. One word or short phrase in reply to teacher elicitation<br>7b. Student silently indicates response<br>7c. Statement that 'expresses judgment, evaluation and the like' (p16).<br>7d. Student can be heard saying something, but can't make it out on the recording<br>7e. Student expresses lack of knowledge<br>7f. Student gives no indication of response | 7a. *It's noun.*<br><br>7b. (Student seems to write or point to the answer - inferred)<br>7c. *I write this, but I don't really think it.*<br><br>7d. (Indistinct)<br>7e. *I don't know*.<br>7f. …. (silence) |
| | Initiate | 8. Elicit | | 8. Student requests verbal response | 8. *But this is noun?* |
| Descriptive labels for further classifying speech | | | a) Correct<br>b) Incorrect<br>c) Korean | a. Student's response to *Elicit* is correct<br>b. Student's response to *Elicit* is incorrect<br>c. Student or Teacher uses Korean | |

To ensure that the data were coded reliably, both intra-rater and inter-rater reliability were measured using the *Coding comparison* function of NVivo 11. Because the data are categorical, kappa coefficient was selected as a measurement of reliability. Intra-rater reliability was calculated by my recoding a random 10% of the conference data (that is, eight transcripts) approximately three weeks after the original coding, and measuring how similar this result was to the original coding; intra-rater reliability calculations resulted in a kappa coefficient of .858, which, according to Altman (1991, 1999: 404), can be considered a very good indication of agreement. Next, inter-rater reliability was calculated with the same 10% of the conference data. The coding scheme was given to a colleague, a doctoral researcher in the applied linguistics research group at the University of Oxford, who agreed to be the inter-rater. Following a brief training session, where we discussed the meanings of the various codes and conducted a trial together, the colleague coded the sample independently. NVivo was used to compare the results of the two raters, providing a kappa coefficient of .749; according to Altman (1991, 1999: 404), this is an indication of a good level of agreement.

## 3.8 Ethical considerations

### 3.8.1 Consent and confidentiality

Ethical approval was received from the Oxford Central University Research Ethics Committee prior to the start of the study. During the first week of the term, students were invited to participate in the study. Their classroom teacher gave them a verbal explanation of the study in English, the script having been prepared by me in advance so that both classes would receive the same information. Students were then each given an information sheet and a consent form, with both documents written in both Korean L1 and English L2. The students were told that the study was investigating CF, but in order to reduce chances of students guessing about which mode of CF delivery they received, they were not informed of the different types and delivery modes of CF that would be used. The students were assured, both verbally (L2) and in writing (L1 and L2) that their participation would have no effect on their classroom marks. They were assured that participation in the study would not result in having to do extra work, but that the study would make use of assignments they would already be doing for their classes. It was explained that they were free to withdraw from the study at any time without penalty. To protect the students' privacy, I later anonymised their work for the analyses.

### 3.8.2 Ensuring no harm to participants

Because two of the interventions were hypothesised to increase grammatical accuracy more than the direct WCF received by the Comparison group, it was important for ethical reasons to ensure that the study did not influence students' grades in the class. In order to ensure that grades for the course were given without consideration of grammar, the teachers created grading rubrics which omitted grammar as a category, focusing instead on the content and organisation of the writing; students' attention was not drawn to the absence of a *Grammar* category in the rubric. I marked only grammar points, and the teachers marked only content and organisation.

The students in the Conferencing intervention group (explained in Section 3.6.6.1) received one-to-one time and attention from the teachers during class, while students in the other two groups worked independently. This means participants in one group received a total of approximately eight extra minutes of focused attention from their teacher over the course of the term. To help eliminate the possibility of some students feeling that the personalised attention for some students was unfair, teachers invited all students to ask questions at the end of lessons, and ensured that their office hours were written on the syllabus; attention was drawn to this orally in class. Students were invited to meet with teachers, or to email teachers with questions if students preferred; it is acknowledged that this could be a weakness in the study, but it was considered necessary for ethical reasons. Participants in PS-I demonstrated relief at hearing that their names were not on the Conferencing group list, indicating that students were generally content not to have to participate this way (*cf.* Section 1.2.2 for possible reasons for this). Both teachers reported in the post-study interview that their students reacted similarly to those in PS-I, and none of the students demonstrated concern about it, so the time given to one group over the course of the term does not seem to have been viewed by the students as an unfair advantage available to only a few students.

This chapter built on the information given in the introduction, providing details about the study. The results of the quantitative analyses of the intervention data will be reported in the next chapter.

# 4 Results of quantitative analyses

## 4.1 Introduction to quantitative results

This study investigated the effectiveness of two modes of CF delivery for improving grammatical accuracy in the L2 writing of Korean first-year natural science majors at a university in South Korea, and compared these effects against those of direct written CF provided to the Comparison group. As explained in the Methodology, I marked all of the grammar corrections selected for focus in this study on all students' work, and the classroom teachers marked the work for content. As discussed in the Literature review, Ferris (2006) has found that teachers instructed to provide indirect feedback on students' work actually had trouble doing so; it was as if they were unable to stop themselves from giving direct feedback. In this study as well, one teacher accidentally marked grammar errors not being considered in the study on the second paragraphs of one entire class (approximately twelve participants) before realising that he should not be doing so. As a result, some (n=3) of the Reflective worksheet Group participants in that class filled in the worksheets for all marked grammar errors, even though they were instructed to focus solely on the numbered grammar errors that had been marked. However, since this teacher's class, like all the classes, contained students from all intervention groups, it can be argued that any effect of the protocol violation was spread across all three groups, with a small number of students in each group being affected, and thus, any effect is likely minimal on the study as a whole. This was confirmed with a mixed ANOVA with *Teacher* as the between-subjects variable; differences in change in error ratios between the two teachers' students were not significant: $F(2,214)=.472$; $p=.624$

Three instances of plagiarism were detected in the pre-test, and these three participants were asked to rewrite the paragraph in the following class while other students were working on their revisions, and the three received feedback in that same class according to intervention group. No further instances of plagiarism were detected during the rest of the term.

### 4.1.1 Research questions to be quantitatively analysed

Differential effects of facilitating feedback use on outcomes

1 What is the effect of feedback delivery mode on:
   a the ratio of all targeted errors per hundred words?
   b the ratio of correct article use to all obligatory article use?
2 What is the effect of facilitating feedback use on the development self-efficacy between initial and final measurements?

Mediating variables which might influence outcomes

3 What are the effects of language analytic ability and initial self-efficacy on:
   a the ratio of all targeted errors per hundred words?
   b the ratio of correct article use to all obligatory article use?

4 For the Reflective worksheet group, what is the effect of the quality of reflection in the worksheets on:
   a the ratio of all targeted errors per hundred words?
   b the ratio of correct article use to all obligatory article use?

## 4.2 Quantitative analyses

In this chapter, the results of the quantitative analyses are presented. As an overview of the students' writing, Table 4-1 below shows the total number of words written and errors made by each group for the three tests (pre-test, post-test, delayed post-test), depicted graphically in Figure 4-1, Figure 4-2, and Figure 4-3.

**Table 4-1 Mean numbers of words written and errors (total and article-only) made for pre-, post-, and delayed post-tests for the three groups**

| | Group | n= | Number of words written by students | | | Number of errors for all target grammar features | | | Number of article-only errors | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Mean | SD | Range | Mean | SD | Range | Mean | SD | Range |
| Pre-test | Conference | 37 | 134.3 | 31.6 | 128 | 7.5 | 3.9 | 14 | 3.4 | 2.2 | 12 |
| | Worksheet | 34 | 156.0 | 68.8 | 314 | 7.6 | 4.2 | 19 | 3.5 | 2.2 | 10 |
| | Comparison | 38 | 133.7 | 42.2 | 206 | 7.9 | 4.4 | 17 | 3.0 | 2.5 | 13 |
| Post-test | Conference | 37 | 130.5 | 27.4 | 124 | 3.7 | 3.2 | 13 | 2.0 | 2.6 | 11 |
| | Worksheet | 34 | 141.4 | 34.1 | 132 | 4.7 | 4.2 | 22 | 2.2 | 2.3 | 10 |
| | Comparison | 38 | 128.2 | 25.4 | 110 | 5.7 | 3.9 | 17 | 3.0 | 2.2 | 10 |
| Delayed post-test | Conference | 37 | 142.5 | 44.6 | 209 | 4.3 | 3.2 | 15 | 1.8 | 1.8 | 7 |
| | Worksheet | 34 | 160.3 | 48.3 | 192 | 4.1 | 3.1 | 12 | 2.3 | 1.9 | 8 |
| | Comparison | 38 | 142.7 | 45.6 | 229 | 5.6 | 3.1 | 12 | 3.1 | 2.1 | 9 |

**Figure 4-1 Total numbers of words written for the pre-, post-, and delayed post-tests for each group**

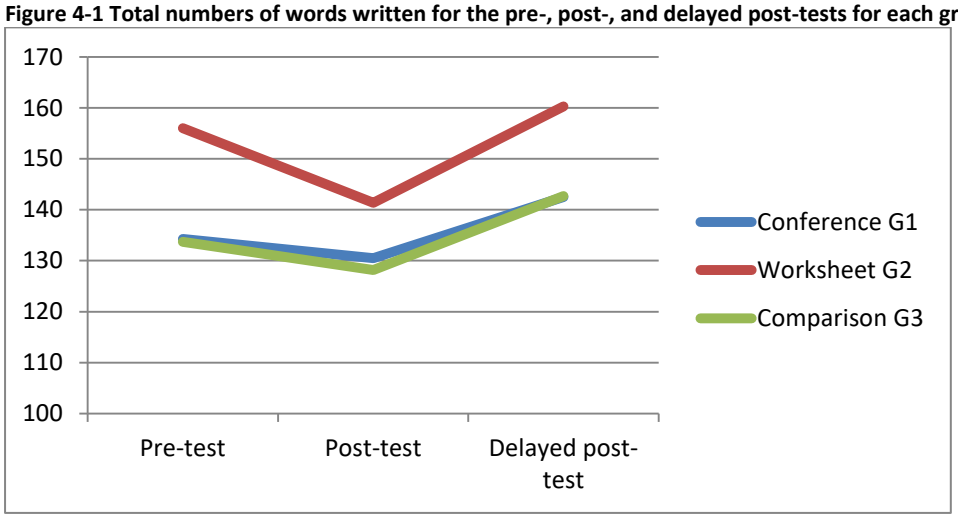



**Figure 4-2 Total numbers of targeted errors made in the pre-, post-, and delayed post-tests for each group**

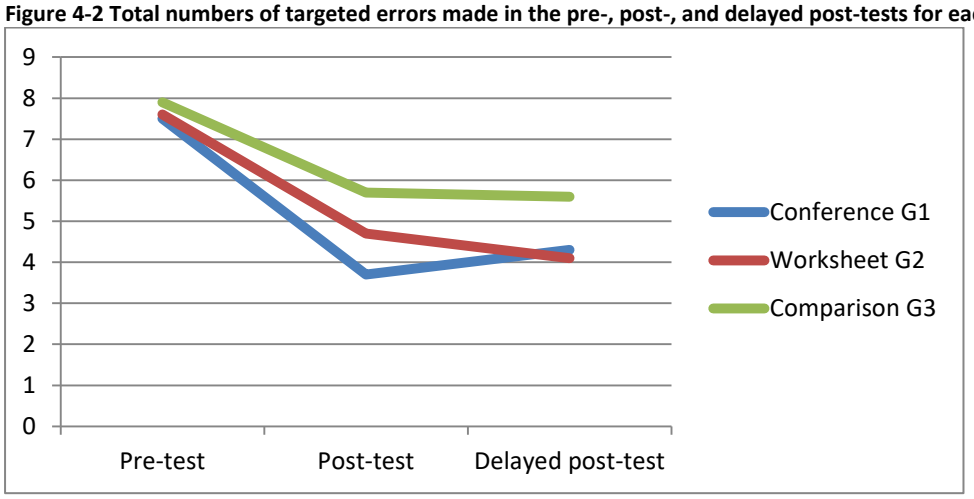
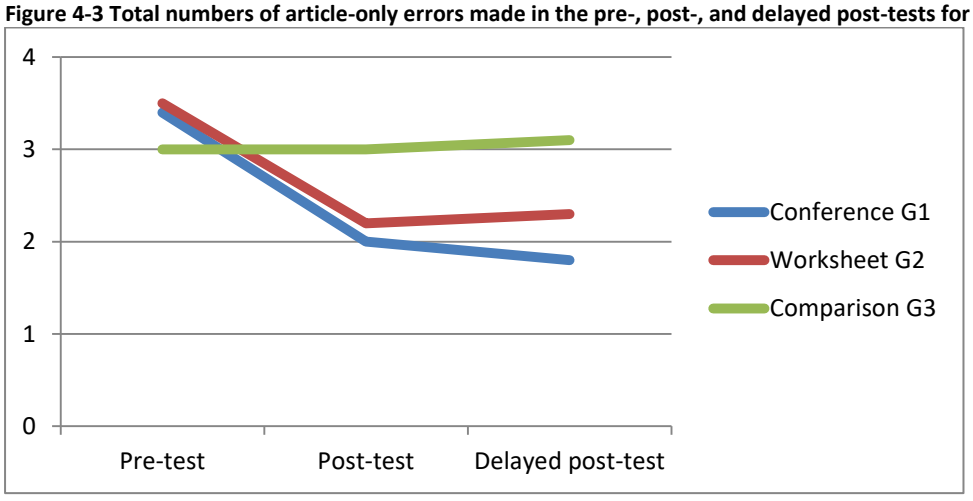



**Figure 4-3 Total numbers of article-only errors made in the pre-, post-, and delayed post-tests for each group**



### 4.2.1 RQ1a Effect of facilitating feedback use on all targeted errors

*What is the effect of feedback delivery mode on the ratio of all targeted errors per hundred words?*

Total words written by each participant in each of the pre-test, post-test, and delayed post-test were counted, and errors in the target grammar features (the five instructed grammatical points and the three points previously taught) were counted; inter-rater agreement was .935, as discussed in Section 3.7.1.1. Using these figures, error ratios per 100 words were calculated and used to examine differences between groups. The second writing assignment was given to provide an opportunity for the second iteration of feedback, as described in Section 3.6.5, and is excluded from this analysis.

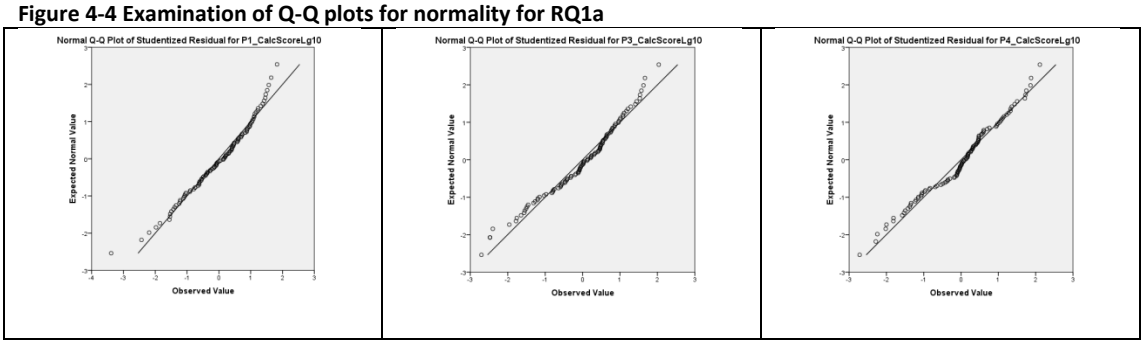
A mixed ANOVA was run in SPSS, version 23, with *time* as the within-subjects factor and *group* as the between-subjects factor on the pre-test, immediate post-test, and delayed post-test scores of all three groups. As the data were found to be moderately skewed, a log10 transformation was applied to the data. Three of the Shapiro-Wilk tests performed on the transformed data suggested that the assumption of normality had been violated (Table 4-2) but an examination of Q-Q plots (Figure 4-4) show that this violation was not severe. Furthermore, as ANOVAs are said to be robust to non-normality (Cochran, 1947), the analysis was continued.

**Table 4-2 Shapiro-Wilk test of normality for RQ1a**

| Log10-transformed error ratios | Intervention group | Shapiro-Wilk | | |
|---|---|---|---|---|
| | | Statistic | df | Sig. |
| Pre-test | Conference | .928 | 37 | .019* |
| | Worksheet | .927 | 34 | .026* |
| | Comparison | .951 | 38 | .097 |
| Post-test | Conference | .966 | 37 | .307 |
| | Worksheet | .981 | 34 | .812 |
| | Comparison | .862 | 38 | .000* |
| Delayed post-test | Conference | .970 | 37 | .400 |
| | Worksheet | .967 | 34 | .396 |
| | Comparison | .957 | 38 | .153 |

*Significant at the .05 level

**Figure 4-4 Examination of Q-Q plots for normality for RQ1a**




Levene's test of assumption of homogeneity of variances was found not to have been violated, as none of the three reports (pre-test, post-test, and delayed post-test) across groups was significant (p=.137; p=.769; p=.089, respectively). There was one outlier (+/- 3.0) as determined by studentised residuals (3.39). The decision was made to keep the outlier as it represented a genuine data point. Box's test of covariance matrices was not significant (p=.075), indicating that this assumption had not been violated. The data were found not to have violated Mauchly's test of sphericity (p=.309).

There was a significant main effect of time, with a statistically significant difference in numbers of errors at the different time points, $F(2, 212) = 37.856$, $p < .001$, partial $\eta^2 = .263$, and an examination of pairwise comparisons showed that these differences occurred between the pre-test and the delayed post-test, indicating that the delayed post-test showed improvement compared to the pre-test. The main effect of group showed that there was a statistically significant overall difference in mean numbers of errors between the three intervention groups $F(2, 106) = 3.767$, $p<.026$, partial $\eta^2 = .066$, and pairwise comparisons revealed that this difference was between the Reflective worksheet group and the Comparison Group. However, no statistically significant interaction was found between time and group, $F(4, 212) = 2.125$, $p = .079$, partial $\eta^2 = .039$. In other words, there were no significant differences between the three groups in the degree to which the ratio of errors had decreased at the time of the delayed post-test.

Paired-samples t-tests were run to confirm whether each group had achieved a significant improvement in accuracy from pre-test to post-test, and from pre-test to delayed post-test. Though significant differences between groups could not be found, all three groups showed significant improvement from pre-test to post-test, and this improvement was shown to be durable, with significant improvement also found between pre-test and delayed post-test scores. As the assumptions are different for t-tests than for ANOVAs, it was unnecessary to transform the data; nonetheless, for consistency, the tests were performed on transformed and untransformed data.

Significant improvement was seen in the results of the tests run on both transformed and non-transformed data. Below Table 4-3 shows the t-test results for pre- and post-test scores and for pre- and delayed post-test scores for the three groups, reported using transformed scores for consistency.

**Table 4-3 T-test results for pre-test to post-test and pre-test to delayed post-test for the three groups**

| | Conference group | Worksheet group | Comparison group |
|---|---|---|---|
| Pre-test – Post-test | t(36) = -5.440, p<.001 | t(33) = -5.451, p<.001 | t(37) = -2.784, p=.008 |
| Pre-test – Delayed posttest | t(36) = -4.742, p<.001 | t(33) = -5.830, p<.001 | t(37) = -3.414, p=.002 |

### 4.2.2 RQ1b Effect of facilitating feedback use on correct article use

*What is the effect of feedback delivery mode on the ratio of correct article use to all obligatory article use?*

Correct and incorrect uses of definite and indefinite articles, within the limits of taught rules, as explained in Section 3.7.1.1, were counted to gauge improvement in article use during the course of the term; the discussion about this focus on articles (as opposed to other grammatical features) is in Section 3.3. An error was classified according to the standard article use required by the context: so if a definite article was required and an indirect article or no article was used, then it was classified as a definite article error, and likewise for indefinite and null article use. To calculate a score for measuring accuracy of use, the number of correct obligatory uses was divided by the total number of obligatory definite and indefinite article uses, resulting in a percentage of correct article uses for each test for each participant. These percentages are provided for each group at the time of each test in Table 4-4.

An issue arose regarding null article use. Some writing samples contained no contexts in which definite or indefinite articles were required, but only contexts in which null article use was required. This is understandable: as the writing was academic in style, writers may have tended toward generalisations, using plural nouns rather than singular nouns with articles. However, Korean is a language without articles. Because null article use can be understood as the default for Korean L1 English learners, correct null article use may be intentional, or it may be a transfer phenomenon. Thus, to create a clear measure of correct article uses, all cases where a null article was required (whether correctly supplied or not) were omitted from the analysis of this research question.

Since this research question therefore focused on the use of definite versus indefinite articles, errors stemming from *a/an* confusion were excluded from the analysis, as using *a* in place of *an* (or vice versa) does not indicate lack of understanding of when to use the definite or indefinite article. In any case, there were only six instances of *a/an* confusion out of 332 article errors.

**Table 4-4 Mean percentage and SD for correct article uses with regard to maximum possible correct uses, by group and time (figures provided for non-transformed data)**

| **Group** | **Pre-test** | | **Post-test** | | **Delayed post-test** | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| Conference (n=35) | 52% | 33% | 73% | 29% | 63% | 37% |
| Worksheet (n=33) | 57% | 32% | 74% | 30% | 63% | 37% |
| Comparison (n=37) | 56% | 34% | 66% | 32% | 57% | 36% |

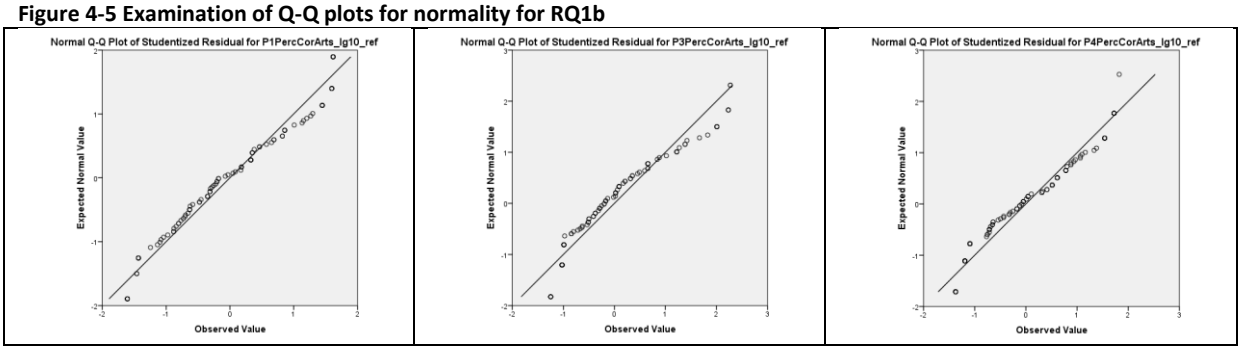
In order to compare change over time in combined definite and indefinite article accuracy for each group, a mixed ANOVA was run in SPSS with *time* as the within-subjects factor and *group* as the between-subjects factor on the pre-test, post-test, and delayed post-test scores of all three groups. There were no outliers ±3.0, as determined by an examination of studentized residuals. Shapiro-Wilk tests suggested that the assumption of normality had been violated (Table 4-5) but an examination of Q-Q plots suggested that these violations were not extreme (Figure 4-5), so the analysis was continued (Cochran, 1947 as mentioned in Section 4.2.1).

**Table 4-5 Shapiro-Wilk test of normality for RQ1b**

| Log10-transformed article ratios | Intervention group | Shapiro-Wilk | | |
|---|---|---|---|---|
| | | Statistic | df | Sig. |
| Pre-test | Conference | .926 | 35 | .022* |
| | Worksheet | .924 | 33 | .023* |
| | Comparison | .919 | 37 | .010* |
| Post-test | Conference | .872 | 35 | .001* |
| | Worksheet | .868 | 33 | .001* |
| | Comparison | .906 | 37 | .004* |
| Delayed post-test | Conference | .869 | 35 | .001* |
| | Worksheet | .906 | 33 | .008* |
| | Comparison | .890 | 37 | .002* |

*Significant at the .05 level

**Figure 4-5 Examination of Q-Q plots for normality for RQ1b**




Levene's test revealed that the assumption of the homogeneity of variances had not been violated for any of the three time points ($p > .05$). Box's test of Equality of Covariance Matrices was shown not to have been violated (p=.849), and Mauchly's test of sphericity also revealed no violations (p=.127).

There was found to be a main effect for time, F(2, 204) = 7.326; p<.001, partial $\eta^2$ = .067, and an examination of pairwise comparisons revealed that this difference was between pre-test and post-test, and between pre-test and delayed post-test. There was no significant difference between the post-test and the delayed post-test, indicating that the improvement in performance was sustained over time. There was no main effect for group, F(2, 102) = .548; p=.580, partial $\eta^2$ = .011, indicating that all groups produced similar numbers of total errors. Additionally, there was no statistically significant interaction between Group and time, F(4, 204) = .572; p=.683, partial $\eta^2$ = .011, indicating that the three groups performed equally well in increasing their accuracy in the use of definite and indefinite articles, and in sustaining this improvement.

### 4.2.3 RQ2 Effect of facilitating feedback use on self-efficacy development

*What is the effect of feedback delivery mode on the development self-efficacy between initial and final measurements?*

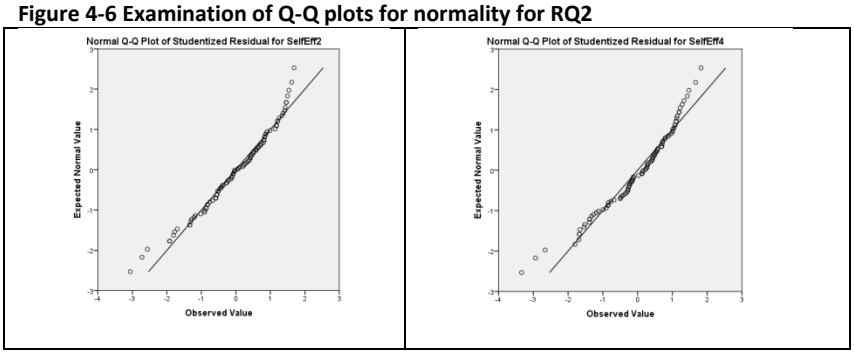
To measure whether there was an effect of the intervention on the development of self-efficacy over time, the difference between the initial and the final self-efficacy tests was used to show change over time. Table 4-6 provides an overview of the self-efficacy scores for the three groups.

**Table 4-6 Mean self-efficacy scores for the initial and final administrations and the gain scores for each group**

| | Initial administration | | | Final administration | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range |
| Conference (n=36) | 73.4 | 11.9 | 45 | 73.8 | 12.5 | 48 |
| Worksheet (n=34) | 70.6 | 13.8 | 56 | 73.9 | 16.1 | 68 |
| Comparison (n=36) | 70.4 | 13.0 | 58 | 74.5 | 15.5 | 66 |

A mixed ANOVA was run in SPSS with *Intervention group* as the independent variable and scores from the initial and final administrations of the self-efficacy measure as the dependent variables. Table 4-6 provides the descriptive statistics for the two administrations of the scale. An examination of studentized residuals revealed two outliers; as they were a) genuine data points, and b) not far outside of the +/- 3.0 cut off range (3.06 and 3.34), the decision was made to keep these data points. Shapiro-Wilk test revealed two non-normal test results, but the Q-Q plots (Figure 4-6) suggest that these violations are minor (Cochran, 1947 as mentioned in Section 4.2.1).

**Figure 4-6 Examination of Q-Q plots for normality for RQ2**




Levene's test of homogeneity of variances revealed that this assumption had not been violated, nor had the assumption of equality of covariance matrices, as measured by Box's test. There was a main effect for time, $F(1,103)=7.213$, $p=.008$, partial $\eta^2 = .065$, suggesting that there was a small improvement in self-efficacy for the participants overall. There was also a main effect for group, $F(2,103)=.136$, $p<.873$, partial $\eta^2 = .003$, suggesting that there were differences in self-efficacy between the three groups. However, there was no statistically significant interaction between Group and Time, $F(2, 103) = 1.170$; $p=.314$, partial $\eta^2 = .022$.

### 4.2.4 RQ3a: Predictive value of LAA and SE for all targeted errors

*How effectively do language analytic ability and initial self-efficacy predict any increase in accuracy from pre-test to post-test of the ratio of all targeted errors per hundred words?*

Table 4-7 provides the descriptive statistics for language analytic ability and initial self-efficacy scores for each group. Although the initial self-efficacy scores were also reported in the previous section, they are included here for clarity.

**Table 4-7 Language analytic ability and self-efficacy scores (mean, standard deviation, and range)**

| Group | LAA | | | SE | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range |
| Conference | 46.3 | 15.2 | 69 | 73.4 | 11.9 | 45 |
| Worksheet | 51.7 | 16.3 | 69 | 70.6 | 13.8 | 56 |
| Comparison | 47.3 | 13.5 | 54 | 70.4 | 13.0 | 58 |

A standard multiple linear regression was run to predict any increase in accuracy from pre-test to post-test of the ratio of all targeted errors per hundred words (dependent variable) from self-efficacy and language aptitude (independent variables). There was independence of residuals, as shown by the Durbin-Watson test (1.827), and there was linearity, as per an examination of scatterplots. The scatterplot also suggested that the assumption of homoscedasticity was met, and, seeing that no independent variable correlations exceeded .7, the assumption of multicollinearity was also met. There were no outliers, as evidenced by the absence of the casewise table that SPSS automatically provides for this information when applicable. There were no leverage points above .2, and no Cook's Distance above 1.0, so these assumptions were met. The assumption of normality was confirmed by an examination of both a histogram and by P-P plots. However, neither language aptitude nor self-efficacy predicted an increase in grammatical accuracy for overall targeted errors, $F(2,102)=.251$, $p=.778$; adj. $R^2$ =-.015. Regression coefficients and standard errors can be found in Table 4-8. A follow-up multiple regression confirmed that neither self-efficacy nor language aptitude predicted improvement by intervention group.

**Table 4-8 Regression coefficients and standard errors for RQ3a**

| Variable | B | $SE^B$ | β | Sig |
|---|---|---|---|---|
| Intercept | 3.016 | 2.117 | | .157 |
| Language aptitude | -.118 | .170 | -.069 | .487 |
| Self-efficacy | -.002 | .026 | -.009 | .928 |

### 4.2.5 RQ3b: Predictive value of LAA and SE on correct article use

*How effectively do language analytic ability and initial self-efficacy predict any increase of the ratio of correct article use to all obligatory article use?*

A standard multiple linear regression was run to predict any increase in accuracy from pre-test to post-test of the ratio of article errors (dependent variable) from self-efficacy and language aptitude (independent variables). There was independence of residuals, as shown by the Durbin-Watson test (1.769), and there was linearity, as per an examination of scatterplots. The scatterplot also suggested that the assumption of homoscedasticity was met, and, seeing that no independent variable correlations exceeded .7, the assumption of multicollinearity was also met. There were no outliers, as evidenced by the absence of the casewise table that SPSS automatically provides for this information. There were no leverage points above .2, and no Cook's Distance above 1.0, so these assumptions were met. The assumption of normality was confirmed by an examination of both a histogram and by P-P plots. However, neither language analytic ability nor self-efficacy significantly predicted improvement in accuracy in the use of articles, as defined in this study, $F(2,99)=1.568$, $p=.214$; adj. $R^2$ =.001.

**Table 4-9 Regression coefficients and standard errors for RQ3b**

| Variable | B | $SE^B$ | β | Sig |
|---|---|---|---|---|
| Intercept | .467 | .257 | | .073 |
| Language aptitude | -.036 | .021 | -.168 | .094 |
| Self-efficacy | -.001 | .003 | -.041 | .679 |

However, given the discussion about the Language aptitude test in Section 3.6.3.1.3, it is prudent to take into consideration for RQs 3a and 3b that despite the students' reported enjoyment of the test, it may have been overly difficult for them; a different test might have provided different results.

### 4.2.6 RQ4a: Effect of worksheet quality on all targeted errors

*For the Reflective worksheet group, what is the effect of the quality of reflection in the worksheets on the ratio of all targeted errors per hundred words?*

This research question focused solely on the Reflective worksheet group. The process of measuring worksheet quality was explained in Section 3.7.1.2; an inter-rater marked 17% (n=6 out of 34) of the worksheets independently, resulting in an inter-rater agreement score of .926. The quality scores from the second worksheet (from both teachers' students) were used for the analysis of this research question, as it was thought that, as the students had completed the worksheet once before and were familiar with it, this second attempt would offer the best measure of languaging. Using linear regression, it was investigated whether the independent variable, the quality score of the second worksheet, could predict the dependent variable, improvement in accuracy of overall grammatical features, as measured by pre-test and post-test gain scores. Table 4-10 provides the mean Reflective worksheets scores for the first and second round of feedback, along with the standard deviations.

**Table 4-10 Mean worksheet quality percentage scores by iteration and by teacher (with standard deviation and range)**

| | Teacher 1's students (n=19) | | | Teacher 2's students (n=15) | | | Both teachers' students | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range | Mean | SD | Range |
| Wkst 1 | 74.3 | 22.0 | 78 | 65.1 | 22.5 | 83 | 70.2 | 23.4 | 83 |
| Wkst 2 | 74.4 | 25.2 | 81 | 71 | 27.0 | 83 | 72.8 | 25.6 | 83 |

Independence of errors was confirmed with the Durbin-Watson statistic (1.545). A scatterplot revealed that there was an approximately linear relationship, and that the assumption of homoscedasticity had not been violated. An inspection of correlations showed no values over .7, suggesting that the assumption of multicollinearity had not been violated. There were no outliers

±3.0, and Tolerance values were in the normal range (greater than 0.1). The independent variable did not predict increased accuracy in all grammatical features, F(1, 32) = .396, p=.534. Running the test again with Log10 transformed data produced similar non-significant results, as did applying the mean worksheet score rather than just the second worksheet score.

### 4.2.7 RQ4b: Effect of worksheet quality on correct article use

*For the Reflective worksheet group, what is the effect of the quality of reflection in the worksheets on the ratio of correct article use to all obligatory article use?*

As in RQ4a, this research question focused solely on the Reflective worksheet group. Using linear regression, it was investigated whether the independent variable, the quality score of the second worksheet, could predict the dependent variable, improvement in accuracy of targeted articles between the pre-test and post-test.

Independence of errors was confirmed with the Durbin-Watson statistic (2.025). A scatterplot revealed that there was an approximately linear relationship, and that the assumption of homoscedasticity had not been violated. An inspection of correlations showed no values over .7, suggesting that the assumption of multicollinearity had not been violated. There were no outliers ± 3.0, and Tolerance values were in the normal range (greater than 0.1). The worksheet quality did not predict increased accuracy in article use F(3, 31) = .218, p=.883.

## 4.3 Chapter summary

This chapter has provided the results of the research questions which could be answered quantitatively (RQs 1-4). For RQ1, the results show that the Conference group, Reflective worksheet group, and Comparison group all improved grammatical accuracy of targeted structures significantly from pre-test to post-test, and that there were no significant differences between the improvement of the groups. The analysis of RQ2 demonstrates that learners' self-efficacy was unaffected by either engaging in dialogue with the teacher to gain a better understanding of their errors or by being prompted to reflect on their errors individually – all three groups maintained similar levels of self-efficacy from beginning to end of the study. RQ3 analyses demonstrated that neither initial self-efficacy nor language aptitude predicted improvement in grammatical accuracy. RQ4 analyses suggest that regardless of the quality of metalinguistic reflections, as measured by worksheet quality (Section 3.7.1.2), improvement in accuracy of targeted structures was similar for all Reflective worksheet group participants.

While this chapter has provided a quantitative, numbers-only view of the results of the interventions, it is necessary to 'look behind the curtain' in order to understand what happened in the study. Chapter 5 will take a close look at what happened in the interventions in order to identify any factors which could help explain the results – some rather unexpected – seen here.

# 5 Results of post-hoc analyses

## 5.1 Introduction

This chapter will present the results of the post-hoc analyses of the interventions. As discussed in Section 2.4.2.2, sometimes interventions fail to proceed as planned. The risk seems particularly high when interventions are implemented by someone other than the researcher, as observed by Ferris (2006), who found that teachers instructed and trained to provide only indirect feedback on students' work had trouble doing so. Similarly, in the current study, while the interventions under my direct control (the Reflective worksheet group and the Comparison group) were implemented with strong fidelity of implementation (FoI), as evidenced by the students' work on their reflective worksheets, an analysis of the recordings of the conferences carried out by the teachers revealed weak FoI. For example, out of 37 first iterations of conferences, none demonstrate complete FoI, and only three came close to following the procedure described in Section 3.6.6.1. Of the 33 second conference recordings, only two recordings demonstrate FoI. This was a particularly surprising finding, given the steps taken to enhance fidelity of implementation (Section 3.6.6.4). The evidence of the lack of FoI will be presented in this chapter.

A possible explanation for the lack of FoI is terminology. *Conferences* seems to be a term commonly used among teaching practitioners, and the teachers in this study would have frequently held conferences in some form with their students throughout their time at the institution, as the institution requires that teachers hold office hours for one-to-one consultations with students. The familiarity of the term might have caused the teachers to create incorrect hypotheses about what the intervention comprised, a preconception which blocked new information I was providing about the protocol for the conferences.

To better understand how prior understanding might have influenced the teachers' understanding, consider Wason (1960), whose seminal research findings suggest that when people have a hypothesis, they tend to seek confirmatory evidence to support their hypothesis, and not seek to disconfirm it, a concept which came to be known as *confirmation bias*. In his study, conducted with psychology students (n=29), Wason found that most participants (n=22) sought out only confirmatory evidence rather than seeking disconfirming evidence, which led participants to err. It is not just inexperienced thinkers, such as Wason's university students, who experience confirmation bias. Nissani and Hoefler-Nissani (1992) demonstrated that even experienced scientists are affected by preconceptions. When scientist-participants found that an incorrect formula provided by the researchers resulted in unexpected answers, most of the participants (n=12) relied on ad hoc

explanations and adjustments of measurements to explain differences between their answers and the expected answer, seeking confirmatory evidence on the accuracy of their hypothesised answers.

Given the level of teaching experience of the teachers in the current study, the voluntary nature of their participation, and the care which they exhibited in other aspects of the study, the type of error in thinking described above, rather than inability or unwillingness, seems the most plausible explanation for the disparity between what they were asked to do and what they actually did. Thus, it is useful to examine the data for evidence of contributing factors to the observed FoI. To accomplish this, post-hoc research questions were formed. These were:

6 How did the two teachers implement the conference intervention?

7 How did students participate in the conference intervention?

8 How did students participate in the worksheet intervention?

9 What, if any, comparisons can be drawn between the two interventions?

The following sections will provide the answers to the post-hoc RQs. Section 5.2 focuses on the analyses of the conference recordings. Next, in Section 5.3, students' participation in the worksheet intervention is explored (RQ8). Finally, an examination of the participation of students in the two interventions is conducted in Section 5.4 in order to what, if any, comparisons can be drawn between the two interventions (RQ9).

## 5.2 Conferences

### 5.2.1 What was the nature of the teachers' contributions to the conferences?

The next sections will explain the results of the data analyses, with a focus on the teachers' contributions to the conferences. Quantification is used only to understand where the problems with FoI arose (i.e. in the protocol), and to illustrate striking findings. The analysis procedure was explained in Section 3.7.3. The aim is to provide answers to RQ6 (how the teachers implemented the conferences). First, Section 5.2.1 will present the analyses of the teachers' contributions to the conferences (RQ6), highlighting where there were deviations from protocol. Section 5.2.2, provides the analyses of the students' participation in the conferences (RQ7), focusing on how this might have contributed to the teachers' lack of fidelity of implementation of the conference protocol. From this, an understanding of the lack of FoI will begin to emerge.

#### *5.2.1.1 Quantitative description of conferences: teachers' contributions*

Before delving into the details of the conferences, it is useful to get a quantitative overview of how the intervention was implemented by the two teachers in order to 'to situate and structure the data' (Grbich, 2013:190). This information will also provide a basis for demonstrating the lack of FoI that will be discussed in detail in Section 5.2.1.2.

Although the teachers had been asked to keep the length of the conferences between four and five minutes, there was wide variation in the actual times. For the first iteration of conferences, the shortest conference was 1.7 minutes long, and the longest was 10.9 minutes long. For the second iteration of conferences, the shortest conference was 1.4 minutes long, and the longest was 10.7 minutes long; this will be further examined below. Four audio recordings from Teacher 2 were lost due to technical difficulties, resulting in the discrepancy between the number of first and second conferences for that teacher. Table 5-1 shows the means, standard deviations, and range of the length of the conferences.

**Table 5-1 Mean length in minutes of conferences (with standard deviation and range)**

| | Teacher 1 | | | Teacher 2 | | | Both teachers | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 7.9 | 1.8 | 6.8 | 3.6 | 1.6 | 4.5 | 5.9 | 2.8 | 9.2 |
| Conference 2 | 5.1 | 2.4 | 9.3 | 2.9 | 1.0 | 3.9 | 4.3 | 2.2 | 9.3 |

Teacher 1: n(umber of conferences)=20; Teacher 2: Conference 1, n=17; Conference 2, n=13

The number of words spoken was selected for facilitating cross-intervention comparison between the conference and the worksheet interventions, with the goal of later comparing student participation in the conferences with student participation on the worksheets (measured by how many words they had written). To compare how much the two teachers were talking in the conferences, it was first necessary to normalise the words they spoke in each conference to control for the different lengths of the conferences. This was done by calculating the total number of words spoken by a teacher in a given conference and dividing it by the total number of errors targeted for discussion in the conference, resulting in the number of words spoken per error (WPE). This measure was selected because it allows both a) to compare teacher and student participation in the conferences, and b) to compare words spoken per error in this intervention with words written per error in the worksheet intervention. Although similarities between speech and written communication can be described, speech is generally less physically demanding than writing (Grabowski, 1996), so it could be argued that this comparison is not useful, as words spoken would

surely outnumber words written. For this reason, it might be unexpected to compare words written and words spoken, given the differences between these two modalities, and indeed, research doing so could not be found. Despite this, the decision to compare words in speech to those in text because, as will be seen in Section 5.4, the data demonstrate that in this study, the assumption that words spoken would outnumber words written was shown to be false. Perhaps as a result of their test-focused English education, students wrote more than they spoke. Kormos (2012) discusses possible causes and effects of this, which will be revisited in Chapter 7.

The means, standard deviations, and range of this analysis are reported in Table 5-2. It can be seen that there is a large difference in the WPE means between Teacher 1 and Teacher 2. These variations will be explored in Section 5.2.1.2.

**Table 5-2 Mean words per error spoken by teachers in conferences (with standard deviation and range)**

| | Teacher 1 | | | Teacher 2 | | | Both teachers | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 38.2 | 16.6 | 64.7 | 32.2 | 13.3 | 51.8 | 35.5 | 15.3 | 73.0 |
| Conference 2 | 40.2 | 22.4 | 93.1 | 20.7 | 5.9 | 21.9 | 32.5 | 20.1 | 93.2 |

Teacher 1: n=20; Teacher 2: Conference 1, n=17; Conference 2, n=13

The first part of the two-step conference protocol (Section 3.4.2.1) was for teachers to prompt students to locate their error within a defined area, e.g. *on this line,* and correct it. Below, Table 5-3, Table 5-4, and Table 5-5 show how many of these *initial* prompts were given, compared to how many student errors there were (that is, how many initial prompts were prescribed by the protocol), along with standard deviations and range. Given that students sometimes spotted an error before being prompted (n=11), it seems that teachers frequently fulfilled the first part of the two-step protocol.

**Table 5-3 Both teachers: mean numbers of errors identified for discussion compared to mean numbers of initial prompts given (with standard deviations and range)**

| | Errors identified | | | Prompts given | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 7.5 | 3.9 | 14 | 5.9 | 2.7 | 13 |
| Conference 2 | 6.6 | 4.0 | 13 | 5.6 | 3.2 | 12 |

Teacher 1: n=20; Teacher 2: Conference 1, n=17; Conference 2, n=13

**Table 5-4 Teacher 1: Mean numbers of errors identified for discussion compared to mean numbers of initial prompts given (with standard deviations and range)**

| | Errors identified | | | Prompts given | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 9.1 | 3.9 | 10 | 7.1 | 2.7 | 11 |
| Conference 2 | 6.9 | 4.2 | 13 | 5.8 | 3.4 | 12 |

**Table 5-5 Teacher 2: Mean numbers of errors identified for discussion compared to mean numbers of initial prompts given (with standard deviations and range)**

| | Errors identified | | | Prompts given | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 5.6 | 3.0 | 10 | 4.5 | 2.1 | 8 |
| Conference 2 | 6.3 | 3.8 | 13 | 5.2 | 2.7 | 10 |

After making the correction, the second part of the two-step conference protocol was for the teacher to check students' understanding, eliciting from the student an explanation about the error. Table 5-6, 5-7, and 5-8 show the mean number of errors that had been identified for discussion and the mean number of checks delivered by the teacher, along with the standard deviation and range. This demonstrates that it was in the second half of the two-step protocol for conferences that fidelity of implementation was most severely affected: teachers did not check students' understanding as called for by the protocol. The following sections will investigate this in detail.

**Table 5-6 Both teachers: Mean numbers of student errors identified for discussion compared to mean numbers of teacher's checks of understanding in conferences (with standard deviation and range)**

| | Errors identified | | | Checks of understanding | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 7.5 | 3.9 | 14 | 1.5 | 1.8 | 7 |
| Conference 2 | 6.6 | 4.0 | 13 | 1.3 | 1.9 | 8 |

Teacher 1: n=20; Teacher 2: Conference 1, n=17; Conference 2, n=13

**Table 5-7 Teacher 1: Mean numbers of student errors identified for discussion compared to mean numbers of teacher's checks of understanding in conferences (with standard deviation and range)**

| | Errors identified | | | Checks of understanding | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 9.1 | 3.9 | 10 | 2.6 | 1.9 | 7 |
| Conference 2 | 6.9 | 4.2 | 13 | 2.1 | 2.1 | 8 |

**Table 5-8 Teacher 2: Mean numbers of student errors identified for discussion compared to mean numbers of teacher's checks of understanding in conferences (with standard deviation and range)**

| | Errors identified | | | Checks for understanding | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 5.6 | 3.0 | 10 | 0.12 | 0.33 | 1 |
| Conference 2 | 6.3 | 3.8 | 13 | 0 | .00 | 0 |

#### *5.2.1.2 Qualitative description of conferences from teachers' perspectives*

Continuing with a focus on the teachers, the following section presents the results of the discourse analysis, adding a more in-depth view of the interactions in the conferences to the quantitative overview in the previous section. The codes were provided in Table 3-17, and the following discussion is organised by examining the coding categories in turn, discussing any relevance, or lack thereof, to the explanation of how fidelity of implementation was compromised, and the implications of those findings. Limited quantification is added where it contributes to the overall picture of the data. Details about how the teachers implemented the protocol will emerge, leading to an understanding of how the teachers might have interpreted *conferences*.

Having quantified the answer to the question about what the lack of FoI of the protocol was in the previous section, this section focuses on qualitative description of moves teachers were using in the conferences; quantification is not normally done. The discussion is organised by the coded moves teachers were identified in the analyses as having used.

##### 5.2.1.2.1 Elicit: ErrorQs

Teachers tended to offer clear prompts – as they were asked to do in the protocol – telling students that there was a problem and providing them with information about its general location (e.g. *There's a problem on this line*) and giving the students an opportunity to find it. Out of 431 prompts for errors, there were no clear cases of a student appearing confused about what was required when

a teacher prompted for an error. Even when teachers shortened the prompt, contrary to what was called for in the protocol, this does not appear to have been problematic, as students demonstrated their understanding by responding appropriately. For example, the following student has made two errors of a similar kind close together, which the teacher elicits concurrently. After the student correctly identifies the need for the definite article in the first error location, the teacher elicits the second correction, saying simply, *And?* rather than giving the elicitation in full. It can be seen that the student understood this and answered it as though it had been a full prompt.

C03Conf01

| **Speaker** | **Transcript** | **Relevant codes** |
|---|---|---|
| Teacher | Now there are two. Two here somewhere. Two mistakes. | ErrorQ |
| Student | Article? | Student elicit |
| Teacher | Yeah. Articles. Two articles. | Reply |
| Student | The | Restricted |
| Teacher | Good job – the. The USA. And? | Accept, Reformulate, ErrorQ |
| Student | The? | Restricted |
| Teacher | Good job! | Accept |

Thus, the first part of the two-step protocol appears to have been followed, and where not followed, changes appear to have been unproblematic.

#### 5.2.1.2.2 Evaluate (Accept and Reject)

When evaluating student responses, teachers offered clear negative feedback, which was almost always followed by *Redirection* ('No, look over here'), *Clue* (e.g. 'It's an article problem'), or *Inform*, where the teacher provided the answer, either the grammatically correct form or the explanation of the relevant grammatical rule, depending on what had been asked.

Both teachers were unambiguous and succinct in their positive feedback when students identified an error and/or its correction: *Good!/Yes./Yeah/ Right.* Teacher 2 showed a preference for *Good job!* to just *Good!* in both the first and the second conferences, but following these expressions of acceptance, both teachers tended to a) repeat the students' correct answer as a confirmation: (*Yeah, <u>the</u>.;* or *Good! <u>A</u> confrontation.*); b) ask for more information (*Good! Where?*); c) move on to the next question (*Right. Ok, the next one is….*), or d) occasionally, move to checking students' understanding (*Good! Why?*). When accepting participant responses, the protocol called for teachers to elicit a metalinguistic reflection; when rejecting responses, the protocol was to provide

the answer and elicit a metalinguistic reflection. As was seen, the eliciting of the metalinguistic reflection was the most commonly missed part of the protocol and a cause of the lack of FoI.

#### 5.2.1.2.3 Elicit: UnderstandingQs

There were only a fourth as many checks of student understanding (n=107) as there were error elicitations. As discussed in Section 5.2.1.1, neither teacher checked students' understanding as frequently as called for by the intervention protocol after the identification and correction of each incorrect grammatical structure, but as was seen, one teacher checked students' understanding considerably more often than the other; of the 107 checks of understanding, all but four were from Teacher 1. Below is an example of Teacher 1 checking for understanding.

J56_Conf01

| | | |
|---|---|---|
| Teacher | (Having finished correcting the plural *Koreans*) That's one problem, *Koreans*. This is also a problem: *Most of Koreans*. Why is that a problem? | Accept; ErrorQ |
| Student | Most Koreans | Restricted; Correct |
| Teacher | Yes. Do you know why it should be *Most Koreans*? | Accept; ComprehensionQ |
| Student | Most of… [indistinct]. | Restricted; Indistinct |
| Teacher | Can you remember what the next word is after of? | Elicit further info |
| Student | Uhh, determiner. | Restricted |
| Teacher | Determiner, good. Or a pronoun. But here we don't have that. Good. | Accept |

Teacher 2, on the other hand, rarely checked the students' understanding of errors, and it is unclear, when he did check understanding, in what situations he chose to do so. For example, in the first instance of Teacher 2 checking for understanding, which occurred near the end of the first iteration of conferences, when the student identified and corrected the first error, the teacher confirmed that the answer was correct and immediately moved on to the next error. There, too, the student correctly identified the error, but this time the teacher asked whether the student understood the reason. Both errors were article errors (the first needing the indefinite article, the second needing the null article), but only the second was checked for understanding.

CQ1 ADH

| | | |
|---|---|---|
| Teacher | Two mistakes are on this line somewhere in here. | ErrorQ |
| Student | (No verbal response; student probably points.) | Silent reply |
| Teacher | Yes, good. What goes there? | More info |
| Student | A | Restricted; Correct |
| Teacher | Yes, *a* vehicle. Good. One more. | Accept; Reformulate; ErrorQ |
| Student | This (probably pointing) | Restricted |
| Teacher | Yes. This you don't need. Do you know why? | Accept; ComprehensionQ |

Five more errors in the same conference are identified and corrected, including one more null article error, but none of the remaining corrections were checked for understanding. If there were a clear pattern, it would be possible to draw some conclusions about Teacher 2's beliefs about and methods of eliciting evidence of student understanding of errors, but none was discernible.

When *Accept* is not followed by a prompt to check for understanding, it can signal the end of the exchange. One of the weaknesses of the IRF framework, as described in previously, is that it fails to engage the students: the teacher asks a question to which there is a correct answer; the student responds; and the teacher accepts or rejects that response. The result of the omission of the prompt for checking understanding is that not only did teachers fail to follow protocol, but they effectively, if inadvertently, discouraged the students from contributing further.

#### 5.2.1.2.4 Elicit: More info(rmation)

When students offered incomplete answers to the teachers' elicitations, teachers often prompted students for more information to complete the answer. They did this without offering clues, simply asking – as can be seen in the example above (*What goes there?*) – for the rest of the information. There were no instances of a teacher offering the answer to a student who had provided a partial answer. It can be seen that teachers were striving to scaffold students' participation in the conferences even though this situation was not accounted for in the existing conference protocol.

#### 5.2.1.2.5 Reformulate

When students' understanding was checked, and when students provided correct answers to them, teachers often reformulated the students' answers. An examination of the reformulations suggests

that they generally two had main functions. One was to make the given answer more complete, such as in the example below:

CL9HJS

| | | |
|---|---|---|
| Teacher | All right, one more (grammatical error) here. | ErrorQ |
| Student | [Silence] | No response |
| Teacher | Start by reading the first part of the sentence. | Narrowing focus |
| Student | - s | Restricted; Correct |
| Teacher | Ah, good job. Why? | Accept; ComprehensionQ |
| Student | It. | Restricted; Correct |
| Teacher | Yeah, *It examines*. Good job! | Accept; Reformulate |

The other common function was to bring the students' answers more in line with the conceptual framework that had been used in the teaching of the relevant grammatical rule. In the example below, Teacher 1 uses keywords that had been used in his teaching of definite articles (for *specific things*) and null articles (for *things in general or generally*).

JK0 JSY

| | | |
|---|---|---|
| Teacher | Good. Rivers and forests. Good. Why? | Accept; ComprehensionQ |
| Student | Many. | Restricted; Correct |
| Teacher | Many, yes. We're talking about them generally. | Accept; Reformulate |

The students also seemed to draw upon the conceptual frameworks that had been used in their lessons. In the excerpt below, the student uses the keyword *Specific*, and the teacher emphasises that there is *only one* of the thing being referred to. The student repeats this information, seemingly to himself, possibly as a way of remembering it.

JK0 KGM_C2

| | | |
|---|---|---|
| Teacher | And there's one more mistake in the same line. | ErrorQ |
| Student | Uhh, a! | Restricted; Incorrect |
| Teacher | No. | Reject |
| Student | The! | Restricted; Correct |
| Teacher | Yes, why? | Accept; ComprehensionQ |
| Student | Specific. | Restricted |
| Teacher | Yeah, there's only one. | Accept; Reformulate |
| Student | Only one. There is one. | Restricted |

#### 5.2.1.2.6 Redirect

When students located errors which were not being targeted for correction, such as spelling errors, teachers redirected them. Usually, this was a two-part process. The first part acknowledged the students' answer: *Well, yes, that's true*… while the second part refocused the student's attention: …*but look at this word here.* No instances of teachers addressing non-targeted grammar were found, so this is evidence of positive FoI.

#### 5.2.1.2.7 Clues

The above examples illustrate students' correct answers to teachers' prompts. However, students did not always answer correctly. The teachers had been instructed only to ask the student once about each error or rule. When students were unable to identify the grammatical error or provide the relevant rule, however, the teachers seemed to want to give the students as many chances as possible to succeed, giving them clues (n=666). This is particularly true of Teacher 1 (n=445), whose longest conference time for the first round of what were to be four-to-five minute conferences was, as reported at the beginning of the chapter, almost eleven minutes. Indeed, examining all conference transcripts revealed that much conference time seemed to be dedicated to giving the students clues about the answers. This occasionally resulted in rather humorous exchanges, as students tried valiantly to cooperate with the teacher even when they did not understand the teacher's clues. This is illustrated in the example below, where the teacher is trying to elicit that the student's phrase, *the emergency situations*, does not need the definite article: the teacher draws on the conceptual framework from the lesson to scaffold the student and help him arrive at the correct answer, though in this case it seems that the student's final answer might have been a guess.

JK0 LJH

| | | |
|---|---|---|
| Teacher | Can you find one mistake in there? | ErrorQ |
| Student | [Silence ] (Seems to be looking in the wrong location) | No response |
| Teacher | No, right around here, toward the end of the line. | Redirect |
| Student | [Silence] | No response |
| Teacher | Have you found the mistake? | Re-elicit |
| Student | No. | Uncertainty |
| Teacher | Emergency situations. Do you mean a specific group of emergency situations or just generally, emergency situations? | Direct reference |
| Student | [Silence ] | No response |
| Teacher | When you say *emergency situations*, do you mean all of them, generally? | Direct reference |
| Student | Maybe? | Restricted |

| | | |
|---|---|---|
| Teacher | You're not talking about a specific situation. You are talking generally about emergency situations. So think about this word. (probably pointing to *the*) | Direct reference |
| Student | The … the | Restricted |
| Teacher | What would I say about coffee? *I ~ ~ Mmmm!* (noisily pretends to take a sip of coffee) | Indirect reference |
| Student | Delicious? | Restricted |
| Teacher | I would say, 'I like …' | Indirect reference |
| Student | Coffee? | Restricted |
| Teacher | Coffee. Generally. | Accept; Offers explanation |

It is interesting to consider why teachers chose this course of action over the conference protocol. Perhaps they felt that giving the student only one chance to make the correction was insufficient, and that if they asked the right question, then the student would be able to answer it. It is also interesting that more clues were given in some situations than in others; a pattern was not found, but it is possible that the teachers were responding to students' body language, which was not captured by the audio recording. For example, if the student wordlessly looked at the teacher rather than at the paper, the teacher might have taken this as a cue to guide the student more specifically. The timing of the intervention might also be relevant; the students had been studying with the teacher for approximately eight weeks at the start of the intervention; thus, it is likely that the teachers had an intuition about the students' degrees of knowledge and ability, and were drawing on this awareness in their interactions with the students.

The clues provided in the conferences can be divided into three main categories:

- Clues which referred the student to relevant parts of their sentence after the error had been identified, such as in the example above where the teacher tries to help the student see for himself that he means *all emergency situations in general*, were coded as ***Direct reference***. This was the most common type of clue provided by both teachers (n=466). Further examination of this category revealed that there were two kinds of *Direct reference:* one kind of *Direct reference*, totaling less than a third of the category, made a specific connection between the error and the conceptual framework which had been discussed in class when the grammar point had been taught, like the example above. These tended to use key phrases from class: *Is it specific or general?* (to help the student think through definite versus null articles); or: *Remember 'most plus noun', but 'most OF plus not a noun',*

*something like the or a or they or them. Remember that? So what should this be?* The second kind of *Direct reference* focused the student's attention to specific parts of the sentence, e.g. *So what's wrong with this verb?* or *What will you put there – the or a?* Clues referring to conceptual frameworks used for explanations were made a sub-category of *Direct Reference* rather than a separate category to avoid creating problems for the inter-rater, who would not have this insider knowledge of the lessons.

- Clues which relied on unrelated examples, such as the coffee example above, were coded as ***Indirect reference.*** This was the least common type of clue (n=56), and it was mostly used by Teacher 1. An examination of the transcripts revealed that Teacher 2 used *Indirect reference* just twice in the study, and both of these instances referred the student to a similar error that had been covered earlier in the student's paragraph.
- ***Clues narrowing focus*** were given in order to help students locate the error. These clues tried to orient the student to the location of the error, such as in the example above where the teacher focuses the student's attention more directly: 'No, right around here, toward the end of the line'. Other examples include providing grammatical hints, rather than geographic ones, such as, 'You're looking for a noun'. This was the second most common type of provided clue (n=144), and for both teachers, grammatical clues were more common than location ones.

Because clues were not included in the prescribed conference protocol but were spontaneously added by both teachers, it is interesting to examine them separately. Table 5-9 shows the total number of errors being treated in the conferences and the total number of clues offered. Following this, Table 5-10 then provides the mean number of clues per error, derived by dividing the total number of clues offered by the total number of errors treated, suggesting that the two teachers offered clues at a similar rate. The clues are discussed in greater detail below.

**Table 5-9 Total numbers of errors made by Conference Group participants compared to number of clues offered by the two teachers in the first and second conference**

| | Number of errors treated | | Number of clues offered | |
|---|---|---|---|---|
| | Teacher 1 | Teacher 2 | Teacher 1 | Teacher 2 |
| Conference 1 | 182 | 95 | 233 | 115 |
| Conference 2 | 137 | 107 | 147 | 72 |

**Table 5-10 Mean clues offered per error by both teachers for both conferences**

| | Conference 1 | Conference 2 |
|---|---|---|
| Teacher 1 | 1.3 | 1.2 |
| Teacher 2 | 1.1 | 0.7 |

The two teachers had somewhat different approaches to providing *clues narrowing focus*. Teacher 1 showed a preference for location clues.

Teacher 1 J56LYB

| | | |
|---|---|---|
| Teacher | The next line has one problem. | ErrorQ |
| Student | (Silence) | No response |
| Teacher | Try looking around here. | Narrowing focus |
| Student | (Silence) | No response |
| Teacher | Ok, how many voice problems are there? | Direct reference |
| Student | (Indistinct) | Indistinct |
| Teacher | Good. So…? | Accept; elicit further info |
| Student | Problem*sss* | Restricted; correct |
| Teacher | Good. Problems is countable, so we'll change that, but pollution is not countable, so this one is okay. And the next problem - we'll skip down to this line. | Accept; Teacher provides explanation; ErrorQ |

In the example above, Teacher 1 elicits an error, to which the student does not respond. The teacher gives a clue narrowing the focus for the error, and again, the student does not respond. The teacher then asks a question which helps identify the error and helps the student understand what the problem is. The response is indistinct, but from Teacher 1's acceptance, it can be assumed that the student responded *Many* or *Two* or something similar. Having established the problematic word and indicated that the problem has something to do with number, the teacher then prompts the student for the correction, which the student correctly provides. The protocol had requested that the teacher use this opportunity to check students' comprehension, but instead, Teacher 1 offers an explanation about why an *s* needs to be added to *problem*, and then, perhaps considering the potential for the student to be confused by the lack of *s* on a nearby noun, adds the reason why *pollution* does not receive an *s*. Without a pause, he then moves to the next problem.

Teacher 2, on the other hand, provided more clues related to parts of speech.

| | | |
|---|---|---|
| Teacher | There are two mistakes. I want you to try to find two mistakes. The first mistake is on this line. Something here. Can you find it? | ErrorQ |
| Student | *Three kinds*? | Restricted; Incorrect |
| Teacher | No, *three kinds of* is good. | Reject |
| Student | Hmmm. | No response |
| Teacher | It's going to be an article mistake in here | Narrowing focus |
| Student | No *the*? | Restricted; Correct |
| Teacher | Right, you don't need *the* because it's just general blood donation. It's not *the donation*; it's *blood donation* generally. That's why it's the same here - the whole thing, you just don't need it. One more on this line here. | Accept; Teacher offers explanation; ErrorQ |
| Student | *The*? | Restricted; Incorrect |
| Teacher | No, it's not an article. This time it's plurals. | Reject; Narrowing focus |
| Student | *Platelets*? | Restricted; Correct |
| Teacher | Ooh, good job. That's right. Not just one – they're really, really small. | Accept; Teacher offers explanations |

In the above example, the student makes a guess following the teacher's initial prompt. The teacher rejects this without further comment, and the student seems unable to progress. Teacher 2 then gives a part-of-speech clue, and seems to point to where the article problem might be. Perhaps seeing that there is already an article in the place indicated, the student suggests that its removal might be the answer. Teacher 2 accepts this answer, and, like Teacher 1, above, offers an explanation instead of prompting the student to supply one, as specified by the conference protocol. As was seen before, Teacher 2 employs key words (*general*) to prompt the student to remember the grammar lesson. Similarly to Teacher 1, Teacher 2 then immediately moves to the next point, where the process is repeated, with the student giving an incorrect answer, and the teacher rejecting it and offering a part of speech clue. The student finds and corrects the error, to which the teacher offers an explanation (*Not just one*).

The above sections describe the various ways teachers tried to engage students and responded to students' answers to their prompts. Teachers followed the prescribed conference protocol initially, prompting students to identify and correct errors on a given line. Correct answers were met with unambiguous affirmation, as the teachers strove to encourage their often-reticent students. If students identified the error but not its correction, teachers prompted students for more information. The teachers seemed to feel that it was important for the students to have ample opportunity to engage with the teachers about the errors; if students were unable to locate the error or correct it, rather than provide the answer and elicit the relevant rule from students, as specified by the protocol, teachers gave students various kinds of clues, leading the student to the answer though hints that narrowed the search for the error, through examples which offered prototypes of the error, or through prompting the student to consider relevant parts of the sentence. Teachers seemed to be scaffolding students with clues, altering the quantity and specificity of the clues according to teachers' perception of student need. After corrections were completed, sometimes the students' understanding was confirmed, and sometimes not, with no clear pattern emerging from the analysis. It is possible that the teachers were taking cues evolving from discussions mediated by the dynamic frame of reference as a way to judge evidence of understanding.

### 5.2.2 What is the nature of students' participation in the conferences?

Turning the focus now to the students, the next section first presents a quantitative overview, which is followed by the results of the discourse analysis. The examination will focus on how students participated in the conferences in order to answer RQ7 (the nature of their participation).

#### *5.2.2.1 Quantitative description of conferences from students' perspectives*

To provide a quantitative overview of the amount of talking students did in the conferences, Table 5-11, 5-12, and 5-13 provide the means for student words per error (WPE; the total number of words spoken by each student in each conference divided by the total number of targeted errors), and for the number of turns taken by students in each conference; Dörnyei and Kormos (2000) also use *Words* and *Turns* as indicators of student participation. The tables here also provide the mean longest run of speech by each student; that is, the number of words spoken in one turn, which adds to the overall picture of how much students were contributing to the interaction with the teachers. For example, by looking at the column describing Teacher 1's students, it can be seen that students took a mean number of 15.2 turns speaking, and the mean number of words spoken together was

3.5; it may be the case that this low level of input from students contributed to the teachers' difficulty implementing the conference protocol. The next sections will qualitatively explore the students' contributions.

**Table 5-11 Teacher 1's students: Calculated words per error, turns taken, and mean longest run of words spoken by students in 1st and 2nd conferences (with standard deviation and range)**

| | WPE | | | Turns | | | Longest run | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 2.9 | 2.1 | 9.2 | 15.2 | 7.6 | 33.0 | 3.5 | 2.2 | 9.0 |
| Conference 2 | 3.7 | 3.3 | 14.2 | 12.6 | 12.5 | 61 | 3.6 | 2.1 | 8 |

Teacher 1: n=20; Teacher 2: Conference 1, n=17; Conference 2, n=13

**Table 5-12 Teacher 2's students: Calculated words per error, turns taken, and mean longest run of words spoken by students in 1st and 2nd conferences (with standard deviation and range)**

| | WPE | | | Turns | | | Longest run | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 2.0 | 1.3 | 5.4 | 9.0 | 6.9 | 27 | 1.9 | 1.7 | 7 |
| Conference 2 | 1.8 | 1.0 | 3.5 | 7.6 | 4.2 | 13 | 2.5 | 1.5 | 4 |

**Table 5-13 Both teachers' students: Calculated words per error, turns taken, and mean longest run of words spoken by students in 1st and 2nd conferences (with standard deviation and range)**

| | WPE | | | Turns | | | Longest run | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range | Mean | SD | Range |
| Conference 1 | 2.5 | 1.8 | 9.27 | 12.4 | 7.9 | 36 | 2.8 | 2.1 | 9 |
| Conference 2 | 3.0 | 2.8 | 14.2 | 10.6 | 10.6 | 61 | 3.1 | 1.9 | 8 |

#### *5.2.2.2 Qualitative description of conferences from students' perspectives*

This section will provide detail and description to the results of the quantitative overview above. As before, the discussion is organised by examining the relevant coding categories in turn, with limited quantification given where it contributes to the picture of student participation.

Despite the differences between the two teachers in mean length of conferences (Section 5.2.1.1) and in the number of speaking turns taken by students, the levels of student participation, as measured by WPE, were low in both teachers' classes, as shown above. The following sections will offer a closer look at how the students interacted with the teachers during the one-to-one student-teacher conferences. While this process of coding student speech clearly required subjective

decisions to be made, the inter-rater and intra-rater reliability scores suggest that the subjective decisions were made with reasonable consistency (kappa coefficients of .749 and .858, respectively, as discussed in Section 3.7.3.4).

#### 5.2.2.2.1 Restricted or silent responses

There were 504 instances of restricted responses, and an additional 159 silent responses (out of 728 total responses), where it could be clearly inferred that students were silently indicating their responses. These silent responses were distinguished from non-answers based on the teachers' comments that followed them; when the recording was silent after a teacher elicited information from a student, and this silence was followed by the teaching providing an *Evaluate*-type response, e.g. *Good, right*., or a request for more information (*So how can you change that?*) or a check for understanding (*Okay, why?*) it was clear that an answer of some kind had been given. On the other hand, when the silence was followed by a *Clue* that functioned to elicit the same information that had already been requested, then it seems legitimate to infer that no answer had been given at all. Silent responses may indicate that the students felt uncomfortable or put on the spot, or perhaps it was a pragmatic move; if the question could be answered without speaking, then there was no reason to do differently.

#### 5.2.2.2.2 Expanded responses

As described above, restricted responses, silent responses, and silences represent the greatest portion of student interaction for both classes. There were only 65 instances of student speech coded as *Expanded* (43 from Teacher 1's students); of these 65 instances, 52 were sub-coded as *Clarification* (37 from Teacher 1's students). Much of the text sub-coded as *Clarification* consisted of students repeating what the teacher had said, e.g.:

JK0 JSY

| | | |
|---|---|---|
| Teacher | How many roads are there? | Direct reference |
| Student | Road? | Clarification |
| Teacher | Yes. How many are there? | Direct reference |
| Student | One | Restricted |

It must be said that one cannot be certain that the student was indeed asking for clarification; the student might have been stalling for time or thinking out loud. However, as it is impossible to know

for certain, and as the teachers seemed to treat these repetitions as requests for clarification (as above), it seemed reasonable to classify them as requests for clarification.

Given previous experience with this population, the planned conference protocol did not account for situations where students might challenge or question the teachers. The data reveal that this omission was not necessarily a poor decision in this context; with both teachers and in both conferences, there were only five instances of students challenging teachers' answers. The example below contains the most student-tutor interaction of these five; it is an example of genuine dialogue, with the student engaging with the teacher about an issue, and persisting until he achieves understanding. Teacher 2 is helping the student correct the sentence *Iodo is the important island*. Iodo is a small Korean island of maritime importance; the teacher uses two other important Korean islands, Jeju-do and Dokdo, to demonstrate why the article should be indefinite rather than definite. The student challenges the teacher's explanations until the student appears satisfied. While it is possible that the topic, being local geography, contributed to the students' active participation, as this topic was also addressed by other students who did not exhibit such active participation, it is more likely that this was the result of individual difference, such as personality or comfort with non-Koreans; a likely explanation is that it is the result of a combination of topic and individual factors.

CN6 HGG

| | | |
|---|---|---|
| Student | A very important | Restricted; Correct |
| Teacher | Right, it's a very important island. Good job. | Accept; Reformulate |
| Student | It's the very? | Challenge |
| Teacher | Not *the* because there are many islands. | Teacher reply |
| Student | If s is here? | Challenge |
| Teacher | No, there are many islands. Unless you say the MOST important, then that's okay. But Korea has many islands: there's Jeju, there's— | Teacher reply |
| Student | But here is one. | Challenge |
| Teacher | Yes, one of many. | Teacher reply |
| Student | Here is specific island. *The* is not correct? | Challenge |
| Teacher | It's not. Because is Jeju an important island? Jeju is important, right? And Dokdo is important, right? So that's three different islands | Teacher reply |
| Student | Ahh | No reply |
| Teacher | So it's *a very important island*. If you want to make it specific, then you would say, *The most important island*, and then there's *the*. Okay? | Teacher reply |

There were only two instances of students giving more information than absolutely necessary in their answers. The longest of these was a student expressing shock or disbelief when told that *pollutant* was a countable noun. This student seems to have a rather outgoing personality, as he is heard making small jokes and interacting with the tutor throughout the conference, unlike students in other conferences.

JK0 KGM

| | | |
|---|---|---|
| Teacher | But there is one more in that line. | Redirect |
| Student | One more? … S? No! | More than necessary; Correct |
| Teacher | Yes. | Accept |
| Student | Yes??? (gasp) It countable? | More than necessary |
| Teacher | Yes, pollution is uncountable, but pollutant is countable. | Teacher reply; explanation |
| Student | Wow. | More than necessary |
| Teacher | So many different chemicals. They're all countable. | Teacher reply; explanation |
| Student | Really! | More than necessary |
| Teacher | Ok, turn over the page, please. Look at line two. | Elicit-ErrorQ |

#### 5.2.2.2.3 Student non-responses

There were 291 instances of students remaining silent following a teacher's attempt to elicit information from them, where there was no evidence of a non-verbal response; the distribution of these between teachers and conferences is shown in Table 5-14. The higher rate of non-response from Teacher 1's students could simply be because Teacher 1 gave more clues, which provided more opportunity to fail to respond. There were also an additional 59 utterances which will be called *audible non-verbal responses* (ANVRs). ANVRs were utterances of ambiguous meaning, such as *uhhh*, *ahhh*, *uhmm*. Sometimes it was easier to guess the meaning than at other times: *Hmm?* with a rising intonation seemed to indicate a lack of hearing or understanding. Without the benefit of video, it is impossible to ascertain whether, for example, *ahhh* meant *Yes! I understand* or perhaps *I need time to think* or even *I feel embarrassed because I don't understand*. These were classified separately in an *ANVR* node rather than in the *No reply* node due to their ambiguous nature, even though, like non-responses, they did not appear to contribute to the exchange with the tutor.

**Table 5-14 Numbers of non-responses given by the two teachers' students in the conferences**

| | Teacher 1's students | Teacher 2's students |
|---|---|---|
| Conference 1 | 136 | 47 |
| Conference 2 | 79 | 29 |

A word tree analysis was run in NVivo in order to better understand what preceded and followed students' silences. Predictably, attempts to prompt students to respond to something, (e.g. *Elicit* or *Clue*) preceded student silences, and these elicitations seemed to be divided fairly evenly between statements and question forms (*Ok, there's a problem on line 2* vs *Ok, do you see problem on line 2?*). Equally predictably, student silences were followed by clues, sometimes in increasing levels of specificity. If students did not know the answer, they seemed to wait until the teacher delivered a question they were able to answer rather than actively seek help from the teacher; there were only ten instances of students telling the teacher that they did not know the answer, and all ten of these were Teacher 1's students. If students were, for some reason, expecting clues to be given, they might have waited for the clues rather than answer the prompt, which could have influenced teachers' deviations from protocol. The excerpt below illustrates how teachers dealt with silences from students.

| | | |
|---|---|---|
| Teacher | The third one is here, let me find it. It's right here somewhere. | ErrorQ |
| Student | Silence | Silence |
| Teacher | Again we're looking for an article so start at the beginning of the sentence. | Narrowing focus |
| Student | Silence | No response |
| Teacher | Where is your noun? | Narrowing focus |
| Student | Where is noun? | Clarification |
| Teacher | Yeah, where? | Teacher reply |
| Student | (Points) | Silent response |
| Teacher | Yeah, it's safety standards. | Accept |
| Student | Silence | No response |
| Teacher | Are these general standards or specific standards | Direct reference |
| Student | Specific? | Restricted |
| Teacher | Yeah, specific to the W.H.O. So what would go here? | Direct reference |
| Student | The! | Restricted; Correct |
| Teacher | Good job - *the*. The safety standards | Accept |

As seen above, students received increasingly specific questions in response to their silence, as the teacher worked toward a question that had fewer possible answers. In the example above, the student was silent when given a broad elicitation. Even when the teacher narrows the focus to articles, the student remains silent. Finally, the student shows willingness to reply, albeit silently, to the request to identify a noun. The teacher confirms that the student indicated the correct noun, and then seems to wait for the student to apply the previous clues to this new information (i.e. the need for an article) and identify the error. Again, the student is silent, to which the teacher asks the student a question with only two choices – general or specific. The student chooses one of the proffered options, and as the teacher begins a further elicitation, the student volunteers the correct answer.

#### 5.2.2.2.4 Elicit (student)

A study discussed in Section 3.4.2.1 reported that students observed that an advantage of conferences was the opportunity to ask questions and request clarification, but there were only three instances of students in the current study eliciting information from teachers. One student, upon identifying the error, asked his teacher for help correcting it. As the students are STEM majors, it is understandable that they might rely on terminology common to their field of study (here, *solve an error* rather than *correct an error*). This could also be the result of L1 transfer, where one word (풀다 (to solve, resolve, answer) is used to talk about the resolution of many different kinds of problems.

J42ChChS

| | | |
|---|---|---|
| Teacher | Good, yes. | Accept |
| Student | Solve, uh….? | Elicit (student) |
| Teacher | Ok, there are two ways you can solve it. You can take this away, or you can add *s*, here, as you can see. | Inform |
| Student | Thank you | Accept (student) |

In the other two cases, one involved a student confirming that he had understood correctly, and in the other the student was hypothesising about other contexts in which the answer could be different. The student interrupted as the teacher began to move to the next point. Note that this is the same student who expressed shock over an answer earlier.

JK0C2 KGM

| | | |
|---|---|---|
| Teacher | And the next line has-- | ErrorQ |

| | | |
|---|---|---|
| Student | Teacher, if… uh.. problem… only problem… *The*? | Elicit |
| Teacher | Right. But you mean this one, big Chinese pollution problem. So there's only one, so it's *the*. | Offering explanation |
| Student | Ahh, *the*. | Accept (student) |

##### 5.2.2.2.5 Korean

The conferences took place in English, with the occasional exception of single Korean words used here and there, usually by students. Two of the seven instances of Korean were by teachers who used single Korean words for emphasis. Of the five used by students, three were parts of speech; when one considers that English education in schools is delivered in Korean, it is perhaps surprising that there were not more instances of this.

#### *5.2.2.3 Summary*

To summarise the examination of student participation in the conferences, students seem not to have taken advantage of the opportunity for individualised attention through one-to-one conferences. It is possible that the apparent reticence was the result of feeling uncomfortable or put on the spot. It is also possible that participants felt linguistically restrained, feeling that they should use only English with their non-Korean teacher. The next section will, *inter alia*, investigate these hypotheses by examining how the Worksheet Group participants engaged with their intervention.

## 5.3 Reflective worksheets

This section reports the results of the analysis of the worksheets, addressing post-hoc RQ 8. First, the worksheet instrument is briefly described; a full description was provided in the Methodology. Next, participation in the worksheet intervention is quantitatively examined. Following that, the worksheet data is qualitatively examined. Finally, in order to facilitate comparison of student participation in the Worksheet condition with student participation in the Conference condition, the coding structure used for the conferences is adapted and applied to the worksheets and the results are reported.

Figure 5-1 shows the top three rows of the reflective worksheet; numbered rows continue down the page, with instructions for students to copy their grammatical errors and the corrections, which had been marked and numbered on their essays, in the corresponding spaces, and to write an

explanation for each. (Note that in Korea, noughts (0) are often used for indicating acceptability and crosses (X) for unacceptability.) In the example, the two noughts provide the same information, demonstrating that both English and Korean answers are acceptable. A note in parentheses then explains why the examples are acceptable. A negative example is indicated with a cross, and for emphasis, it has a line drawn through it. An explanation of why this is an unacceptable answer is given in parentheses.

**Figure 5-1 Worksheet example**

| | Error | Correction | Explanation of the correction in 한글 or English (or mixing the languages) |
|---|---|---|---|
| ex. | Yesterday I go | go→ went | O: Use past tense verbs to talk about past events<br>O: 지났던 일에 대해서 말할 때 과거형 동사 사용함<br>(Both examples above explain the rule)<br>X: ~~'went' is better~~ (This doesn't explain the rule) |
| 1. | | | |

### 5.3.1 What was the nature of student participation with the worksheets?

The following sections will examine the data resulting from the discourse analysis of the worksheets in order to address RQ8 (participation in the worksheet condition). First, the process of adapting the conference coding structure to the worksheets will be explained. Applying a similar coding structure as was used with the conferences ensures that similar information is extracted from the data of the two interventions. This consistency will also facilitate comparison between student participation in the two interventions. Next a quantitative overview is provided. This is followed by the details emerging from the discourse analysis.

#### *5.3.1.1 Adapting the coding structure to the worksheet data*

Firstly, all *Tutor* codes were removed, since the only input to students was direct written corrective feedback and there was no tutor input. Next, the *Student* codes were addressed. Changes were made to *Qualification* of students' responses, as the responses could no longer be captured with a simple *Correct/ Incorrect* dichotomy; whereas partially correct answers in the conferences had been completed through dialogue, students in the Reflective worksheet Group received no such scaffolding. Since the worksheet quality scores accounted for accuracy and completeness, worksheet quality scores were used as a measure of correctness. Table 5-15 summarises how the Conference coding scheme was applied to the Worksheets.

**Table 5-15 Adapting the Conference coding scheme to the Worksheet analysis**

| Conference codes | Worksheet codes | Reasons for adaptation |
|---|---|---|
| *Tutor(all codes)* | (Removed) | Tutor did not participate in the worksheets |
| Silent reply | (Removed) | All replies were silently written, so it is more useful to examine Restricted vs Expanded responses. |
| Correct/ Incorrect; | Worksheet quality scores used instead | This was adapted to account for partial correctness |
| Restricted/ Expanded response | Expanded response | Restricted answers were expected, given the small space; only expanded answers were flagged |
| Uncertainty | Uncertainty | (No adaptation) |
| No response | No response | (No adaptation) |
| Indistinct | (Removed) | The handwriting on many of the worksheets was poor, but it could eventually be decoded by someone accustomed to Korean script written with little care. |
| Elicit | Elicit | (No adaptation) |
| Korean | Korean | (No adaptation) |

#### *5.3.1.2 Quantitative overview of student participation with the worksheets*

This section will provide a brief quantitative view of how the students engaged with the reflective worksheets. In order to provide an estimate of the amount of student participation with the reflective worksheets, Table 5-16 provides a summary of the number of words written. The mean numbers of words written by students is similar between iterations of the intervention, though the standard deviations are large.

**Table 5-16 Mean numbers of words written on reflective worksheets for both teachers for both iterations of the intervention (with standard deviation and range)**

| | Teacher 1's students (n=19) | | | Teacher 2's students (n=15) | | | Both teachers' students | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range | Mean | SD | Range |
| Wkst 1 | 42.7 | 25.9 | 77 | 34.7 | 13.1 | 45 | 39.2 | 21.3 | 74 |
| Wkst 2 | 37.3 | 24.5 | 97 | 40.9 | 16.9 | 63 | 38.9 | 21.3 | 95 |

#### *5.3.1.3 Qualitative investigation into student use of worksheets*

This section will provide the details of the results of the qualitative analysis of the worksheets, expounding on the quantitative information provided above. The worksheets are examined in light of the adapted coding structure. What follows is a discussion, organised by examining the coding categories in turn, of the student participation in the intervention.

##### 5.3.1.3.1 Correct/ Incorrect (conceptualised as Worksheet quality)

To summarise the process of measuring worksheet quality, two aspects of the students' self-explanations, accuracy and completeness, were used in marking each answer on a scale of zero to two, as explained in Section 3.7.1.2, in order to derive a worksheet quality score; a percent was then calculated from that score. Table 5-17, also provided in Section 4.2.6, shows the mean percentage scores and standard deviations of worksheet quality scores.

**Table 5-17 Mean worksheet quality percentage scores by iteration and by teacher (with standard deviation and range)**

| | Teacher 1's students (n=19) | | | Teacher 2's students (n=15) | | | Both teachers' students | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Range | Mean | SD | Range | Mean | SD | Range |
| Wkst 1 | 74.3 | 22.0 | 78 | 65.1 | 22.5 | 83 | 70.2 | 23.4 | 83 |
| Wkst 2 | 74.4 | 25.2 | 81 | 71 | 27.0 | 83 | 72.8 | 25.6 | 83 |

These results indicate that students were often able to independently explain their errors. In interviews, students mostly reported being able to find the relevant rules in their notes and textbooks. Even though no feedback was given on the worksheets, the scores remained relatively stable between the first and the second iteration of the intervention, particularly for Teacher 1's students. It is interesting that Teacher 2's students' scores started lower than Teacher 1's students, but that they increased their worksheet quality scores for the second iteration of the intervention; no clear reason for this was found in a comparison of the worksheets from the two iterations. (A similar measure of the frequency of correct answers from Conference group was rejected because the answers were often drawn from the students gradually by the teachers in conversation; therefore, that they were able to eventually offer the correct answer might be considered less meaningful.)

##### 5.3.1.3.2 Expanded responses

The worksheets were examined for instances of *Expanded responses*; that is, instances where students a) challenged the marked error or its correction (no instances); b) requested clarification

about something on their work (no instances); c) offered more than the minimum necessary to answer the question; there was only one instance of a student writing more than the minimum necessary. In this case, the student wrote not only an explanation of the rule, but also examples to go along with each explanation. Interestingly, the student did this only on the second iteration of the intervention – recall that no feedback was given on the reflective worksheets. It is possible that the student had independently decided that adding concrete examples of the grammatical points would help him remember. It is possible that he was drawing on the examples provided in the lessons when these points were taught. When the student was explaining the need for the definite article, he wrote, '*ex. There is a window. The window is clean'*. When he needed to explain the role of the indefinite article, he wrote '*ex. I have a ball'*.

#### 5.3.1.3.3 Uncertainty

Next, the transcripts were examined for expressions of uncertainty, such as *I don't know* or a question mark (*?*) in the space provided for an answer; no instances of these were found. The two most likely reasons for the absence of uncertain responses are that: 1) direct CF had been provided, removing any doubt about what the error was and what its correction needed to be; 2) students were allowed to use their notes and textbooks to find the answers when completing the worksheets. Another possible reason is the students' English education background; test preparation classes and materials, with which students seemed to have extensive experience, advise against leaving questions unanswered (Educational Testing Service, 2016). Thus, it is possible that the students were accustomed to answering written questions even when they lacked certainty about the answer, resulting in them treating the worksheet as a test, filling in all of the answers out of habit. Another possible reason is that even if students lacked knowledge of the relevant rules, they were able to find enough information in their notes or textbooks to write something, even something brief, as an answer.

#### 5.3.1.3.4 No response

Worksheets were then examined for *No response*, of which there had been many in the conferences, when students did not reply to tutors' prompts. There was one instance of a blank worksheet; when asked about it, the student, who happened to be an interviewee, said he had no memory of seeing it before; his data were excluded from statistical analyses. One student filled in his errors and the corresponding corrections, which had been provided via direct CF, in the first two columns of both worksheets, but left all of the spaces for explanation blank; this student was considered not to have completed the intervention, and his data were excluded from statistical analyses. A further six

students had instances of some unlogged errors on their otherwise completed worksheets; four of these appeared to have been occasional errors missed out by accident, but two students had omitted logging three and four errors respectively. The reason for this is unclear; of course these could be the result of accidental oversight, as the omitted errors did not seem to be beyond the students' abilities – they were mostly errors of a similar type as others logged on the students' paper.

There was only one instance where a student clearly appeared not to know the answer and chose to leave the explanation space empty rather than attempting an answer. This case differed from the others because the student successfully logged all of the errors, and successfully explained most of them; the only three which were left unexplained were those where the definite article was needed, and had been incorrectly omitted. It seems possible that the student felt that he lacked the knowledge necessary to explain the need for the definite article. Six of the nine worksheets which contained *No response* codes were from students in Teacher 2's class; the remaining three, two of which were from the student who left the reflective column blank on both worksheets, were from Teacher 1's classes.

The paucity of non-responses, despite feedback not being provided in the worksheet intervention, is interesting. It could merely indicate that students are accustomed to being assigned and completing seatwork, and thus completed the worksheets by force of habit; it could also indicate that they felt the activity could benefit their learning, and thus found value in completing it. As before, there is also the possibility that this is a result of test-preparation classes, where students learn not to leave questions blank, and are thus in the habit of offering a response to all questions.

#### 5.3.1.3.5 Korean

The investigation into language choice when completing the worksheets revealed two notable things. Firstly, with the exception of just two students, all of the worksheet explanations were completed entirely in Korean. There was only one student who chose to complete both iterations of worksheets entirely in English, and one additional student who completed only the first worksheet entirely in English. In all other worksheets, English was used only when it was essential to the explanation. In the example below, the student incorporates the English word *many* in the explanation because of its contribution to the explanation.

N6_KDCh_WS2

| | | | |
|---|---|---|---|
| 1. | many kind | kind → kinds | Many 가 있으니까 복수를 싸용함 |

*(Use plural because of 'many')*

If the English word was not essential to the explanation, most students used Korean-only explanations

C3_JJJ_WS2

| | | | |
|---|---|---|---|
| 3. | produce | produces | 주어가 단수 |

*(Subject is singular)*

This initially surprising finding is less remarkable when the students' English learning background is taken into consideration. As described in Section 1.2.2, and as students confirm in their interviews in Chapter 6, their English education has been test-focused in structure; since the main focus for English study in Korean middle and high schools is to prepare for the multiple-choice university entrance exam, English instruction is given in Korean L1. An exception to this occurs when the class is a speaking or writing class, but because of time constraints, these tend to be restricted to English cram schools specialising in communication skills rather than test preparation.

## 5.4 Comparison of student participation in the two interventions

The following section applies the information from previous sections to address RQ9 (an examination of the two interventions). First, a quantitative overview is given in order to provide an overall picture of how the two interventions compare with regard to student participation. Note again that teacher participation is not considered here, as the teachers did not have a role in the worksheets. Drawing from the results of the prior discourse analyses, a more detailed discussion of qualitative aspects of comparison is also provided.

In order to compare student participation in the two interventions, Table 5-18 provides the mean numbers of words spoken or written (as appropriate) per error.

**Table 5-18 Student participation in the two interventions, as calculated by mean errors made divided by students' words spoken in conferences or written on worksheets (Standard deviation in brackets)**

| | 1st iteration of intervention | | 2nd iteration of intervention | |
|---|---|---|---|---|
| | Conference | Worksheet | Conference | Worksheet |
| Teacher 1's students (n=19) | 2.9 (2.1) | 5.6 (2.9) | 3.7 (3.3) | 4.4 (3.3) |
| Teacher 2's students (n=15) | 2.0 (1.3) | 6.1 (2.2) | 1.8 (1.0) | 6.6 (4.7) |

By this measure, levels of student participation were often at least twice as high on the worksheets as in the conferences. This difference must be interpreted cautiously, as writing and speaking are quite different actions which might not be easily compared (Section 5.2.1.1). Nonetheless, one might expect that more words would be spoken than written, so the finding is unexpected. It is possible that students in the Reflective worksheet group felt more relaxed than those in the Conference Group, who might have felt put on the spot. Students in the Reflective worksheet group also had time to search in their notes and books for answers, and to answer the questions in any order, perhaps saving those they found to be more difficult until the end. In addition, they were working with a familiar pedagogical tool – pencil and paper – while the Conference group participants were experiencing a new (to them) pedagogical tool. Even though the Conference participants had an expert facilitating their use of CF, helping them make direct links between the sentences on their paper and the conceptual frameworks used in the classroom when teaching the grammar, it is possible that feelings of discomfort were strong enough to nullify any advantages the conferences might otherwise have offered. The extent of this difference between students' previous educational experiences and this experience will emerge from the analyses of the interviews. A last difference is that of language; as discussed next, the Reflective worksheet participants were using L1.

It is interesting how English was used in the two interventions. Though there are only three examples (produced by two students) of worksheets containing all-English metalinguistic explanations, the explanations on all three of these were written using multiword answers. This contrasts with what was seen in the conferences in Section 5.2.2, where students tended to offer, also in English L2, only minimal responses. Compare the two excerpts below, which focus on explaining the same kind of error. The first is from one of the two students who offered English explanations on the worksheets; the second is from a conference student.

Worksheet participant 62_BDG_WS1

| | | | |
|---|---|---|---|
| 1. | then hit… | hit → hits | Front 'makes', so it also 'hits' because of the subject 'it'. |

Conference participant: CL9HJS

| | | |
|---|---|---|
| Student | - s | Restricted; Correct |
| Teacher | Ah, good job. Why? | Accept; ComprehensionQ |
| Student | It. | Restricted; Correct |
| Teacher | Yeah, *It examines*. Good job! | Accept; Reformulate |

This differences between the groups, as seen in Table 5-18 and in the examples above, might be explained by the relative lack of pressure on the Reflective worksheet group; they had time to review their notes and to compose their thoughts in English L2, and to use tools to scaffold their own learning, while the participants in the Conference group might have felt pressure to answer quickly.

## 5.5 Chapter summary

This chapter has reviewed the Conference Group and Worksheet Group interventions. Qualitative analysis results indicate that the lack of fidelity of implementation (FoI) in the conferences tended to originate from a) teachers neglecting to elicit an explanation of the relevant grammar rule from the student after the error had been identified, even though this was called for in the conference protocol, thereby terminating the exchange and discouraging students from further engagement in that interaction; and b) teachers offering clues to students in the process of eliciting errors, though the intervention protocol did not call for this. This breach of protocol may have been stimulated in part by what appeared to be student reticence. The results of the analyses conducted on the data from the two interventions suggest that, qualitatively, there were differences between student participation. It could be argued that students seemed to engage more with the Worksheets, which allowed them to use their L1, to make use of external tools such as notes and textbooks, and to take their time completing the task. The implications of these findings will be considered in the Discussion chapter.

# 6 Results of interview analysis

## 6.1 Introduction

The aim of this chapter is to provide answers to the final research question, exploring the perceptions of the student- and teacher-participants about the interventions. The importance of this question – providing triangulation as well as allowing conclusions to be drawn about the pedagogical viability of the CF interventions from the perspective of the participants – was discussed in Section 3.4. The method of analysing the interviews was described in Section 3.7.2. The first section discusses student participants, so within Section 6.2, *participants* will refer to the students; Section 6.3 focuses on the teachers, so in that section, *participants* will refer to the teachers. Where confusion is possible, the participants will be referred to as *student*- or *teacher participants*, as appropriate.

## 6.2 Student participant interviews

Following explanations about how the interviews were conducted and some considerations about the analyses, the first section (6.2.1) specifically aims to answer RQ5a: *What are student participants' perceptions of their intervention?* by exploring how participants understood their intervention, and then by examining their experiences with the intervention. The next two sections provide contextualisation for these perceptions. Section 6.2.2 takes one step back to allow those perceptions to be understood within the context of the classroom. The next section (6.2.3) takes a larger step back in order to obtain a nuanced view of the students' perceptions by exploring their prior English learning experiences.

The student participant interviews (n=13; recruitment described in Section 3.6.3.1.5) were held in a café, after finding in the first pilot study (PS-I) that the participants seemed intimidated by being in a large empty classroom alone with a non-Korean interlocutor (Section 3.5.1). On a few occasions, the café became noisy and it was sometimes necessary to ask an interviewee to repeat him- or herself. Even so, this trade-off was deemed worthwhile, as even the least talkative interviewees seemed more at ease in this social setting compared to the interviewees in PS-I. The only major disruption that occurred during the interviews was a fire drill, which took place near the end of one of the interviews, but the student and I picked up our belongings and went outside, where we were able to complete the interview, though the ringing of the fire alarm caused some distraction; fortunately, this interviewee was one of the more talkative interviewees, so the only noticeable effect of the interference was some requests for repetition.

Each participant is given a unique three- or four-letter identifier followed by a digit – 1, 2, or 3. The digit indicates which group the participant was in.

1: Conferencing group (n=4)
2: Reflective worksheet group (n=5)
3: Comparison (direct WCF) group (n=4).

Table 6-1 provides an overview of the characteristics of the interviewees: their teacher; their gender; their proficiency; their L1 writing experience (*cf*. Section 3.6.2); their pre-test to post-test overall gain (+) or loss (-) score; their initial self-efficacy score; and their language analytic ability score.

**Table 6-1 Interviewee characteristics**

| Grp | Student | Teacher | Gender | CEFR | L1 writing | Writing gain/ loss[1] | Initial self-efficacy | Language aptitude |
|---|---|---|---|---|---|---|---|---|
| 1 | JSY1 | Teacher 1 | M | B2 | 1-3 | +5.5 | 72% | 15% |
| 1 | LJH1 | Teacher 1 | M | B2 | 1-3 | +6.5 | 78% | 38% |
| 1 | HJS1 | Teacher 2 | F | B2 | 1-3 | +0.4 | 68% | 54% |
| 1 | LHH1 | Teacher 2 | M | B1 | 0 | +5.8 | 78% | 31% |
| 2 | JSL2 | Teacher 1 | F | C1 | 0 | +1.4 | 77% | 69% |
| 2 | HBH2 | Teacher 1 | M | B2 | 0 | -3.6 | 54% | 31% |
| 2 | OJH2 | Teacher 2 | M | B1 | 0 | +1.7 | 78% | 54% |
| 2 | YJH2 | Teacher 2 | M | C1 | 1-3 | +1.8 | 76% | 54% |
| 2 | KYM2 | Teacher 2 | M | B2 | 0 | +5.0 | 82% | 62% |
| 3 | KJH3 | Teacher 1 | M | B1 | 0 | -3.1 | 60% | 54% |
| 3 | JJW3 | Teacher 1 | F | B1 | 1-3 | +2.3 | 89% | 46% |
| 3 | YChH3 | Teacher 2 | M | B1 | 1-3 | +5.0 | 64% | 46% |
| 3 | ChHJ3 | Teacher 1 | F | B1 | 1-3 | -4.8 | 32% | 38% |

[1]A positive number indicates an improvement in error ratio score (i.e. in grammatical accuracy) from pre-test to post-test

A discussion of how traditional 'problems' in interviewing can be beneficial when viewed through the lens of responsive, reflective qualitative inquiry (Roulston, 2011) was provided in Section 3.6.3.1.5. One potential problem when applying responsive, reflective inquiry is that of participant acquiescence (Breakwell, 2012); that is, when participants readily agree to statements made by the researcher regardless of their true thoughts. However, there is evidence in the data that my participation in these interviews matches that called for by Roulston (2011) and is unlikely to have

resulted in acquiescence. Participants seemed more comfortable confirming propositions I offered them than offering their thoughts outright, so I was sometimes forced to verbalise what I thought the students were trying to say and allow them to respond; this was also seen in the Conferences (Section 5.2.2). Potential cultural influences which might explain this will be discussed in Section 7.2.3.2.1. The magnitude of the possible threat of acquiescence is diminished by evidence that the participants were frequently willing to correct me when I incorrectly interpreted what they wished to say. This is illustrated in the example below:

| | |
|---|---|
| Interviewer | When you saw this paper, this is your first essay, and you saw only two problems, were you happy? |
| ChHJ3 | (Shakes head) |
| Interviewer | No? How do you feel when you see just two problems? |
| ChHJ3 | 제가 이거를 쉽게 써가지고. / It's just that I wrote this using simple English. |

In the example above, I guessed that the participant was pleased with her work, as it showed few errors. The participant responded that no, she had, in fact, not been happy with her writing; that is, she was not simply acquiescing by saying *Yes*. Throughout the interviews, participants demonstrated, as here, that while they might have been unwilling or unable to express their ideas independently, they wished to be understood, and they were willing to work with the interviewer to ensure this understanding. This supports Roulston's (2011) claim that flexible, responsive interview techniques can be considered as unproblematic and valid. While it cannot be known whether some participants allowed some misunderstandings to pass without correcting them, the coherence of the content in each interview suggests that if this happened, it was minimal. Regardless, the possibility that there was some degree of undetected acquiescence is acknowledged.

The next section will discuss the interview data. Where participants spoke in Korean L1, the original Korean is provided, along with a translation following a backslash (/). Where the participants spoke in English L2, the transcription is provided in the words used by the participants; if clarification seems necessary, I have added this in parentheses. In order to keep the focus on the experiences and words of the participants themselves, I have, where possible, avoided including my own words in the transcripts except in places where including my words bring clarity to the exchange (Saldaña, 2016:17).

It is important to note that not all participants offered comments on all issues. As an example, in the next section, when asked to describe the nature of the intervention, only two of the four Conference

group interviewees gave answers. Of the two who did not comment on this, one was comparatively uncommunicative on that day and tended to give short, curt answers (LHH1), and the other said he could not remember the manner in which his teacher had provided feedback (HSJ1). Other examples of interviewees not commenting on an issue include cases where interviewees did not understand the question or seemed to feel ill-at-ease. Cases like these reduce the value of quantifying the comments (e.g. *Two of four interviewees said x*); Furthermore, as discussed in Section 3.7.2, what makes a theme is 'not necessarily dependent on quantifiable measures – but rather on whether it captures something important in relation to the overall research question' (Braun and Clarke, 2006:82). Therefore, quantification has generally been avoided.

### 6.2.1 Student participant perceptions of the interventions

The literature indicates that there can be a mismatch between teachers' intentions for and student perceptions of feedback (Hyland, 2013a; Hyland, 2013b). Therefore, student interviewees were prompted to describe their intervention in order to gain insight into what *their* perceptions were about the nature of the interventions. Each section examines participants' perceptions of the 'inner workings' of their interventions.

#### *6.2.1.1 Conferences group interviewees*

When asked to describe how they had been given feedback, the descriptions of the conferences offered by some participants are intriguing. While effort had been made to design conferences which avoided making the participants feel as though they were 'lexico-grammatical games' (Section 2.4.2.2), it was seen in Chapter 5 that conferences were not always carried out in this way. This may be why two Conference group interviewees seem to describe the conferences as guessing games. For example:

> LJH1 [The teacher] says this sentence is a problem. You find. So I find.
>
> JSY1 앞에 와서 뭐가 틀린 것 같냐고…. 한개는 틀렸는데 맞춰봐라 / When I got to the teacher, he said, What looks wrong to you? … There's one thing wrong; guess what it is.

The above descriptions were provided by participants without visible signs of annoyance or frustration, but this finding supports the warning in Hedgcock and Lefkowitz (1996) that learners can perceive guided questioning as riddles or games. The 'guessing game' descriptions here are not necessarily an indication that the interviewees felt the conferences were meaningless; nonetheless, it seems that a teacher or researcher wishing to use conferencing in the classroom would need to

take care to ensure that they were conducted in a way that made it clear to the learners what the purpose was.

Even though student participation in the conferences was seen in Chapter 5 to be low, and even if participants might have perceived the conferences as guessing games, it is still possible that they felt they were able to receive some benefit from them. When asked about his experience receiving explanations directly from his teacher, JSY1 offered an enthusiastic *Yes!* in response.

JSY1 [When I didn't understand], he [the teacher] explained.

Researcher Oh, and could you remember that [explanation]?

JSY1 네! / Yes!

It is possible that the participant wished to please me, the researcher, by giving a positive response, or perhaps hoped that I might give a good report to the teacher. However, given the candid nature of participants' interactions seen thus far, while acknowledging these as possibilities, it seems unnecessary to be overly sceptical about the participant's comments. The other participant describing the conference as a guessing game also indicated that the conference had been useful:

LJH1 [The teacher] told me … uhhhh (gestures a straight line to self)

Researcher Ah, he told you directly?

LJJ1 Yes

Researcher And could you understand his explanation?

LJH1 Oh yes, very easy.

However useful they might have been perceived, the 'guessing game' description of the conferences seen here suggest the possibility that, at least for some participants, the intention of the intervention might not have been clear. Given the discussion in Chapter 5, however, this finding is perhaps not entirely surprising. This issue will be explored further in Chapter 7.

#### *6.2.1.2 Reflective worksheet group interviewees*

The comments from the Reflective worksheet group reflected the intention of the intervention more than the Conference group comments did, and four interviewees described the worksheets as prompting them to recall or review learnt rules. For example, when asked about his feedback, HBH2 explained his view of the reflective worksheet procedure, detailing how it had been useful to him.

HBH2 (My teacher) write this [pointing to direct WCF on paper]. So I think this [pointing to worksheet]. And I rewrite.

Researcher How did you find these corrections and reasons?

HBH2 This (direct WCF) I … see it and remember this and Ah! Ah! [snapping fingers and opening eyes wide, seeming to indicate suddenly remembering something].

The comments from HBH2 seem to confirm that the worksheets might have been useful in activating students' declarative knowledge. JSL2, after explaining the procedure, was even able to recall instances of where the erroneous grammatical form was corrected in later work.

Researcher What did you do when [the teacher] gave you [your work] with the corrections?

JSL2 I found the (relevant handout) or textbook (pages) to change this.

Researcher Okay, do you think the worksheet helped you remember the problems when you wrote (the next assignment)?

JSL2 Yes. [Looking through her work] I don't know where it is directly, but the preposition that I made a mistake was changed (on the next assignment).

JSL2 is seen above to describe how the reflective worksheets helped her use prepositions correctly in the next writing task she was given. However, she also admitted that there was sometimes a disadvantage to working independently. She explained that finding the handouts or textbook pages that related to the identified problem was sometimes challenging.

Researcher Was the worksheet easy for you to do?

JSL2 Not that easy because I can't find that specific situation that I have to use this grammar.

This stands in stark contrast to the interviewees who commented on the ease of understanding the explanations when they were provided in conferences, and she was not alone. As mentioned in Section 3.6.6.3, the teachers were uncomfortable withholding information from their students, so if participants wished to receive explanations, they were invited to ask their teacher for explanations and receive answers to their satisfaction. YJH2 reported taking advantage of the teacher's invitation for help.

YJH2 If I don't understand [the problem], I ask the teacher. And he explain to me.

Researcher All right. Did you ask about some of these (gesturing to the worksheet)?

YJH2 Yes, here, number 3…. Some of them are difficult for me, so I ask professor.

Help-seeking will be further discussed in the next section, and is acknowledged as possibly contributing to the lack of statistical significance between the groups (Section 7.2.1.1).

#### *6.2.1.3 Comparison group interviewees*

Like the Reflective worksheet group interviewees, Comparison group interviewees, receiving direct WCF, also seemed to portray their feedback more closely to how it had been intended than the Conference group interviewees did. Three interviewees in the Comparison group explained that they had received corrections written on their papers, such as the two interviewees below.

ChHJ3 [CF] 그냥 써 주셨어요. / [The teacher] just wrote the correction.

JJW3 [The teacher] change the problem on the paper, and we ask him if I have a problem about the grammar.

JJW3 explained that he had tried to make use of the CF:

JJW3 I remembered these problems in paragraph 1 and I tried to write correct grammar (in paragraph 2)

Interviewees in the Comparison group, as in the Reflective worksheet group, took advantage of the opportunity to ask questions in order to gain understanding about some of their corrections. Two Comparison group interviewees mentioned asking questions, as exemplified by KJH3.

KJH3 [The teacher] tell us read it (the corrected work) and if you don't understand, ask me. So I asked him some question I don't understand here.

Researcher Okay, did you ask about all of them or just one or two?

KJH3 Just one. This. [points to a place on her writing]

Possibly indicating the potential for the interviewees to comprise a biased sample, JJW3 enthusiastically emphasises that for him, shyness was not an issue that prevented him from asking for help.

Researcher Was it easy for you to ask (the teacher) questions about grammar, or did you feel shy?

JJW3 Not shy!

The Comparison group received direct WCF without their understanding of it being facilitated. Unsurprisingly, then, a problem with direct WCF, identified by ChHJ3, was that it was sometimes unclear why corrections were necessary. Unlike YJH2, seen above to show a willingness to interact, ChHJ3 did not take up the teacher's offer to answer questions.

Researcher Was it easy to understand why [the teacher made the correction]?

ChHJ3 No.

Researcher Oh, ok. Here [the teacher] writes *the*. Here, can you understand why?

ChHJ3 (Giggles and shakes head to indicate *No*)

This student seems to exemplify the argument that direct WCF is insufficient because it fails to help learners understand why the correction was necessary. ChHJ3 knows what the error is, and she knows what the correction is; she demonstrates metacognitive awareness that there is a gap in her knowledge, but she does not try to fill that gap. Not prompted to find the answer, she seems not to have sought out the information on her own; nor has she asked the teacher.

Help-seeking by non-Conference group participants did occur, as seen in reports by both the Reflective worksheet and Comparison groups, so it is somewhat surprising that the Conference group participants seemed not to engage more when they were side-by-side with the teacher in a conference given for that purpose, though possibly as a result of individual differences, there was a small number of conferences in which participants actively took part (Section 5.2.2.2.2). The help-seeking seen in the interviews may be a result of bias stemming from the interviewee selection procedure. As mentioned at the beginning of this chapter, interviewees on the whole might be skewed toward the type to participate actively: the interviewees were all volunteers, and only those who demonstrated on the background survey that they were able and willing to elaborate beyond the minimum were invited to attend an interview. Despite possibly being skewed toward more outgoing, engaging participants, the interviews were useful for triangulation, and they produced useful insights about interviewees' current and prior experiences.

### 6.2.2 Student participant perceptions of the class

Section 6.2.1 focused on the interventions. This section widens the focus to allow for a view of what was happening in the classroom, giving context to the interventions. This is valuable because it will consider the participants' views of the study more broadly, rather than just their views of the

intervention part of the study. This section is divided into two parts. Section 6.2.2.1 reports on participants' perceptions of the targeted grammatical forms and instruction. Section 6.2.2.2 details what participants shared about the effort they had put forth in the class.

#### *6.2.2.1 Grammar*

The grammatical structures targeted for instruction were taught throughout the first seven weeks of the term. Perhaps as a result of participants' prior test-focused instruction, in which they were taught a structure and then, instead of using it in language production, were asked to identify its correct form from a list of options, interviewees expressed surprise at the gap they discovered can exist between *understanding* and *use*. This is exemplified by KYM2, who emphasises his view that the grammatical structures being taught in the current class seemed easier than those from his secondary schooling.

KYM2 [The grammar we studied in this class] is more… it's easier than I learnt before this class, so it is… it is no difficult to learn that. …. But small things…! (laughs, indicating articles and prepositions)

While KYM2's comment about 'small things' could be interpreted to mean that he found articles and prepositions, which tend to be short (small) words, challenging. Articles and prepositions are ubiquitous in English, and therefore, it could be expected that students would have had frequent encounters with them. The resulting sense of familiarity with these 'small words' might have caused him to underestimate the difficulty of using them. It is possible that when the teachers presented the lessons, the learners dismissed the information as unessential. Perhaps some awareness-raising activities could have been used to raise learners' awareness of the challenges of using these features (procedural knowledge) as opposed to just knowing about them (declarative knowledge); cf. Section 2.3.1.1. This will be further explored in Chapter 7. Whatever the cause, the view that there is a gap between understanding something and using it is recurrent. Five other participants, such as LJH1, expressed similar sentiments.

LJH1 [The grammar] is easy.

Researcher Was it easy to use or easy to understand or both?

LJH1 Easy to understand.

Researcher How about using it?

LJH1 Ah, uhm…. (laughs, shakes head).

LHH1 offers more detail about this experience, commenting about understanding the grammatical features that were explained, and then discovering when receiving his marked work that it contained errors of things he thought he had known.

> LHH1 이해하는데 적용이 힘들어요….그때 그때는 이해 할 수 있는데, 뭐지, 다음에 쓸 때도 많이 틀렸죠. / I understood [the grammar points], but it was hard to use them…. I understood the explanations when they were given, but, how can I say this, I had many errors on the next writing assignment.

Even though the participants seemed to view the targeted grammatical features as 'easy' ones, as above, compared to grammatical features studied in their secondary school English classes, there were indications that perhaps some of the participants became more conscientious about accuracy in writing.

> YJH2 I am conscious about my error, my mistake. … Although I try not to make an error, sometimes unconscious…..

This awareness that they were making errors with grammatical features they felt they understood seems to have caused them to view the writing tasks as more difficult after the CF was provided than they had felt before the CF. LJH1 voices his view that, having received feedback on the first writing task (and, thus, having his awareness raised), he found the second writing task more difficult. When asked about this, he explains that while he wrote the first task without thinking about the grammar too much, having his awareness raised by the feedback caused him to look at his next writing task and wonder what unseen errors might exist.

> LJH1 (The second writing) is difficult.
>
> Researcher Oh, it was more difficult?
>
> LJH1 First (we wrote) and (then the) second… that time… I don't know where wrong.
>
> Researcher Ok, so when you write the second one, you think, *Oh, where are the problems? I don't know!*
>
> LJH1 Yes.

This seems to indicate that CF was prompting LJH1 to try to notice his errors. If this were common amongst all participants, it could make the case for the benefit of having awareness-raising activities prior to instruction. For example, perhaps if the student had had this experience prior to instruction, he would have benefitted more from the instruction and ensuing writing practice.

When asked how they handled the realisation that the grammatical features they had thought of as easy were actually difficult to use in their writing, one participant demonstrated a refreshingly bold approach to errors and writing. Even after JJW3 realised the difficulty of using the grammatical features she had been taught, she refused to let herself become too focused on the grammar. She reported trying to use correct grammar and pay attention to the feedback she received, but her focus was on the communicative aspect of the activity.

JJW3 영어를 쓰는데 .. 막… 문법이 틀리거나 그런것에 대해 망설이지 않고 그냥 자신있게 영어를 쓰는게…. / I'm writing in English, and, well, I can't hesitate over things like grammar. I just write with confidence.

Researcher So after you do it again and again…

JJW3 And fix!

Upon receiving CF, students reported experiencing a variety of emotions. Some felt dejected or nervous; three, like YJH2, describe their main reaction as one of shock or surprise.

Researcher How did you feel when you received your paper?

YJH2 First, didn't. Just shock.

Researcher A good shock or a bad shock?

YJH2 Bad shock. When I write I think it's right, and then I see there's many grammatical error.

Two interviewees mentioned using strategies from CSAT preparation classes to prepare for writing tests. LHH1 reported reviewing the PowerPoint slides from the lessons, and JSY1, while admitting he did not study at home, explained that his strategy was to try harder to remember the lessons and focus on his writing.

Researcher 준비는 어떻게 했어요? / How did you prepare (for essays)?

LHH1 그냥 그래머 PPT 다시 보고. 그래죠 / I reviewed the (teacher's) PowerPoint slides about grammar. That's about it.

Researcher How did you remember these problems (for the next time)?

JSY1 Not study at home, but I remember these things very, so I write more concentrating.

These excerpts suggest that some first-year university students in the current study might have been struggling to make the transition from viewing English from a test-focused perspective to viewing English as something to be improved communicatively, by use. This, too, as in the above discussion

about interviewees underestimating the difficulty of using the grammatical forms, seems to suggest that some learners may have perceived the instruction on the grammatical features as unessential, despite evidence of struggling to use their knowledge.

Perhaps participants realised too late that their knowledge was incomplete, and then lacked the time needed to master the targeted grammatical features. HBH2 pointed out that this was his first time using English productively, as opposed to being tested on it. He felt that with time and opportunity, he could improve.

HBH2 It's first time write English and grammar, so I think I (will) remember (with) time.

This section has provided some thoughts shared by interviewees. Their comments seem to indicate that the class introduced pedagogies that were likely unfamiliar to the participants. The sections thus far have focused on what was happening in the classroom. The next section looks at what the learners report about their effort outside of the classroom.

#### *6.2.2.2 Effort*

All interviewees who spoke on the topic freely admitted they had not put much effort into the class. This is representative of the larger sample of the current study: 87% of survey respondents (total n=108) reported on an anonymous end-of-term survey that they had spent zero hours (n=67) or less than one hour (n=27) each week studying English. KYM2 spoke of having a desire to study, but finding himself unable to carry out the actual work.

KYM2 Frankly speaking, I don't study except homework and [attend] class.

Researcher Are you too busy, or do you think it's not necessary?

KYM2 I think I have to do, but I don't do that.

Another interviewee gave a similar account: KJH3 had specified on the background survey that one of his reasons for wanting to study English was to experience the culture (given this context, he likely meant American culture (Section 3.6.3.2.2)). Despite having expressed desire for English and having given a motivation for studying it, he was unable to carry out his plan.

Researcher On average, how many hours did you spend studying English every week?

KJH3 Zero.

Researcher If you want to learn English and you want to experience this culture [as the student had noted on the background survey], why did you not study more? Was it motivation?

KJH3 Not time. And… 귀찮아서/ it's a pain, you know? [laughing] Before I came to (university), I, my mind was *Study English!* …. But….

Comparing the background surveys of the study sample to these accounts given by interviewees, it seems the desire to learn was there, but it was not strong enough to overcome all of the hurdles that come with university life.

An example of one of those hurdles was given by OJH2 as *time*; first-year students in this institution have informally reported taking even nine classes, so it is not surprising that students with heavy course loads might be forced to prioritise some studies over others.

Researcher 영어 공부하는게 좀 흥미있다는게 하는데 머 수업시간에만 공부했다는거 좀 어떻게 하게됐어요. 좀 재밋으면 더 많이 할 것 같은데.. / You said that you enjoyed English, and yet you say you didn't study outside of class. I'm curious about that. If you enjoyed English, I'd have thought you'd have studied more….

OJH2 그죠.. 그게, 재미있는거랑.. 재미가 있는데 막 하는거랑 다르더라구요. 어쩔땐 그니까, 배우며는 거부감이 있거나 그런건 없고 …. 하루막 이렇게 투자하기는 힘들드라구요. / Right, about that. Enjoyment…. Enjoying something and actually doing it are different. It's not that I am resistant to studying…. It's hard to find the time in a day to invest in studying.

OHJ2 specified that he enjoyed English, but that enjoyment alone was not a strong enough force to drive him to study. It may be relevant to note that current class was only worth two credits, despite it being a three-hour-per-week class. The participants in the current study were not language majors, so it likely made sense to them to focus their energies on the courses that would 'count' more.

The introduction chapter provided a discussion of the students' learning histories and of the social expectation that their first year of university is a time to celebrate having 'made it'. HJS1 explains that though she came to university with the plan to study English, she gave in to the desire (or peer pressure) to spend time with her friends instead.

Researcher Okay, you said you like English and you want to speak well and you want to write well. I'm curious why you said you didn't study.

HJS1 I planned to study English, but my (classmates) … don't study first year, so I… I play outside (with them).

It would be understandable that these students, who had given up their extra-curricular and social activities for three or more years to achieve the goal of entering a highly ranked university (cf. Section 1.2.2), might relax and focus on non-academic pursuits after having achieved that goal.

Those years of effort to gain admission to a good university could be judged successful, given their entry at the highly ranked institution where the study took place, but LHH1 reported entering university with a sense of hopelessness, as if he already believed English proficiency was beyond his grasp.

LHH1 글세요. 이런거… 동기부여? 동기부여 할 때 좀 하겠죠. / Well, motivation? I guess when I have motivation (for English) then I'll do better at it.

If other participants in the study sample shared this hopelessness, this might help explain some students' unwillingness or inability to force themselves to study, an issue which will be further explored in Section 7.2.3.2.1.

It is understandable that if one believes one has no hope of mastering something, then it could negatively influence one's willingness to try. However, Table 6-1 shows that this interviewee's self-efficacy score is 78%, which is higher than the overall participant mean of 71.5%. As discussed in Section 3.6.3.1.2, self-efficacy focuses on a specific task; in this case, *understanding grammar explanations and using them in writing*. Thus, what initially seems like a large difference between LHH1's words and his initial self-efficacy score may simply be the result of the construct's necessarily narrow focus. This was the participant mentioned in Section 6.2.1 as being comparatively uncommunicative and tending to give curt answers; it is possible that the participant was having a difficult day.

KJH3 and HBH2 also admitted to having given up on English, but for practical, calculated reasons. They seemed pragmatic about where they thought their attentions should be focused:

KJH3 Because I'm good at math, so I study math a lot so … because I can get more score in math, so I study math. English is little boring and too many remember words, so I little bit … little study.

HBH2 Math is… like… math very knowledge. Math is difficult but I get many knowledges. English is difficult (and) I didn't get many knowledges.

HBH2 seems to be indicating his view that while other subjects than English are also difficult, they offer greater benefit. As the learners in this study were not majoring in English, this view might not be surprising.

Even though one participant said she had not understood the CF she had been given, she did not ask the teacher for explanations. When asked why she did not ask for help, her reply seems to reflect a disheartened spirit:

ChHJ3 제가 교수님한테 물어봐도 영어로 말씀하셔가지고 이해를 못 할 것 같아서…. / Even if I were to ask the teacher (about the CF), he'd answer in English, and I doubted I'd have been able to understand anyway.

It is common to see English L2 learners, particularly in some countries in East Asia, who have experienced many years of English L2 education and yet display high discomfort when called upon to use the language for communication (Williams, 2017). Perhaps such students might have been more comfortable with a Korean L1 language teacher. The background survey asked students about their preference for Korean L1 vs. English L1 language teachers; of the 122 survey respondents, 35 (29%) said they preferred Korean L1 teachers. Of these, most (n=31) indicated that ease of communication through Korean L1 (e.g. understanding instruction, asking questions) was the main factor in their choice; the other four respondents indicated just feeling more comfortable with Korean L1 teachers. An examination of the background survey confirms that this interviewee was amongst those preferring a Korean L1 teacher. As a reason for her preference, she wrote on the survey:

> 영어를 잘 하지 못해 외국인 선생님의 말은 못 알아 들을때가 많다. / Because I'm not good at English, there are many times I'm unable to understand what the foreign teacher is saying.

The participant's experience seems to affirm the value of English L2 instructors who can speak the learners' L1; such language teachers have long been the norm in many countries, but in Korea, English (and by extension, English teachers) is still commonly viewed as a foreign commodity to be imported from the West (Williams, 2017:18ff).

When asked how the teacher might have helped her, ChHJ3 seemed not to have any ideas.

> ChHJ3 선생님보다 제가 노력을 해야 됐었는데 / More than something the teachers could do, it was me who needed to have put forth more effort.

While the possibility exists that this is merely a case of a student accepting responsibility for herself, it may be useful to look beyond the surface to seek answers which are more deeply rooted. The next section explores participants' prior learning experiences to see what influencing factors might have affected the participants and their perceptions of the interventions and the class.

### 6.2.3 Student participant perceptions of prior learning experiences

Research studies, including longitudinal ones as the current study, take place at a particular point in time. In order to understand the participants in a given study, it is useful to consider not only their performance in that study, but also their experiences prior to the study which may have influenced them. This section discusses what students shared of their prior English learning experiences. This information is valuable because it will allow for a deeper understanding both of their views of the current class and of their perceptions of the interventions. It can also provide a basis for understanding the seemingly low effort put forth by students. For example, continuing with ChHJ3, the participant discussed at the end of the previous section, she said she had given up on English while still in secondary school, and had consciously made the decision to put her energies elsewhere.

> ChHJ3 제가 영어를 잘 못하다 보니까 대학에 들어올려면 잘하는게 있어야 되잖아요. 그래서 영어 아예 포기하고… 다른 걸 열심히 했었고. 영어를 잘못해서…. / I realised I wasn't good at English, but when you go to university, you have to be good at something. So I completely gave up on English and worked hard at other things, since I was bad at English.

It is unfortunate that students are entering university with such negative perceptions of their ability to learn English. As English is a gateway subject (Section 1.2.2), the potential for negative knock-on effects is great.

A recurring theme in all participants' reporting of their English education was lots of reading and translation work, and they frequently referred to multiple-choice type tasks. They talked about CSAT-focused classes, and all reported that L2 writing had not been a part of their learning. This appears to support the information students provided in their background surveys about their writing experiences. The excerpts also support the description of this population given in the Introduction Chapter as test-focused language learners.

ChHJ3 ‘Middle school and high school… didn’t write English…. Only solve grammar problem’.

JJW3 ‘Well when I was in high school we study English is for 수능 (CSAT). Very boring and solving problems. 수능 (CSAT) doesn’t ask like this (gestures to essays) grammar. I just solved the problem…. Reading and just find the answer.

Researcher ‘그러면 중학교 때, 고들학교 때 글 쓴 적이 있었어요? 영어로?’ / Then in middle school or high school did you ever write anything? In English?

LHH1 ‘없었어요’. / No, never.

The interviewees describe their grammar practice as *solving* grammar problems; perhaps not coincidentally, the language resembles the guessing game descriptions given by some of the conference interviewees. It also seems to resemble the kind of grammar instruction students tend to be given in secondary school, where they are given a sentence with, for example, a gap or an underlined portion, and asked to choose between the options listed what best fills the blank or what would correct the underlined portion. The experiences provided here bear resemblance to reports from students in test-focused environments elsewhere in the region (e.g. Tian and Lowe, 2013; Williams, 2017).

Interviewees explained that they were not asked to produce language creatively in English L2 classes, rather, the focus was on identifying the best answer from those given. The potential of this type of educational background to impact on learners’ ability to participate in conferences will be examined in the Discussion Chapter. JSY1 sums up his English learning experience memorably:

JSY1 ‘I.. frankly, I never use this writing…. Only right thing and wrong thing and select it’.

Interviewees seemed aware of the reason their previous language teachers had focused on multiple-choice-type English instruction, as exemplified by YHJ2 and HBH2, below. If the goal is to pass a multiple-choice test, then the instruction methods could be expected to reflect that purpose.

YHJ2 Most of the English in middle school and high school is reading and test.

Researcher Preparing for 수능 (CSAT)?

YHJ2 Yes. I don’t have chance to write the essay or paragraph.

HBH2 If I can solve the problem (it’s enough). I didn’t (need to) write an essay (for CSAT).

Even though they showed an appreciation for why they needed to master what might be called test-focused English (that is, the ability to solve English questions by choosing the correct answer), students also were aware that English is more than just a test subject. The participant from above continued, explaining the gap between language study for testing and language study for communication.

HBH2 (Speaking of the current writing class): This way is better study.

Researcher When you say *this way*, do you mean directly trying it yourself?

HBH2 Yes, this way is good…. High school English is only 'solve the problem'. Solve the English problem is not related to using English.

Despite years of test-focused English education, HBH2 seems to be indicating that he sees potential for English L2 beyond the classroom. If other interviewees shared this view, it might explain their initial enthusiasm for English seen on the background survey and infrequent carrying out of their plans to study.

One interviewee was exceptional, in that she did report having occasional L2 writing experiences in school, but only for tests. When asked about it, she explained that feedback was not a part of the writing experience. When asked how the teacher had helped her with her writing, she replied that the teacher had just lectured and had not provided individualised help. Thinking that she might have misunderstood the question, I pointed to feedback given on her work in the current class and asked if that teacher had done anything similar. She indicated that feedback was not given on writing.

Researcher Were these tests or just homework?

HJS1 Test.

Researcher So how did the teacher help you?

HJS1 Not help me, just teach.

Researcher Did your teacher do like this (gesturing to feedback written on essays) if you had a grammar problem?

HJS1 No.

HJS1 explains her view that her teachers had not been sources of help, *per se*, but rather sources of information ('Not help me, just teach'). If this were the case, it might indicate that the expected outcome of the assignment, even from the teacher's perspective, was the grade. Grades are important, and it is understandable that students might fixate on the grade received for their work

rather than on what they learn in the process of doing it (Section 3.6.5). However, providing only grades and not feedback seems to suggest a very different conceptualisation of what might be viewed as *teaching* in the West. Students coming from such backgrounds into a Western teacher's class might understandably struggle to understand why the teacher was providing feedback. As HJS1 is a Conference group participant, it could also be understood if she struggled to understand what was expected of the discussions of errors. When asked how she was able to gauge her English ability in her secondary school class, she explained that the grade told her all she needed to know:

Researcher How did the teacher change your grammar problems?

HJS1 Submit and they give me grade.

Researcher So how do you know about your writing?

HJS1 I know after the grade.

Even for learners reporting some L2 writing experience, it seems that the view of writing as a process was indeed foreign to them. If participants with writing experience report not having experience with CF, as above, then it is possible that participants in the larger sample, who indicated very little writing experience, may not have had the opportunity to develop the skills needed to take in feedback, work to understand it, and then use it in future work. This might contribute to explaining the low participation seen in Chapter 5. This will be further explored in Chapter 7.

## 6.3 Teacher participant interviews

This section provides a discussion of what emerged from the interviews with the teachers. The teachers' perceptions are useful because this information allows for a discussion of the interventions from the view of the people implementing the interventions in their classrooms. Section 1.3 proposed that one of the contributions of the current study is an ecologically valid study design; the teachers' perspectives about the interventions can offer an indication about whether the interventions had real-world value (Graham and Harris, 2014).

There were only two teacher participants, which resulted in a less substantial amount of data compared to the student participants. When preparing the teachers' transcripts for analysis, the transcripts were read through several times to understand what the broad themes were. Teachers' comments seemed to focus on two main areas: a) their current experience with this study; and b) feedback methods they used prior to this study.

### 6.3.1 Experience with interventions in the current study

The teachers were most involved in the Conference intervention, as the Reflective worksheets were only seen by me, the researcher. The experiences the teachers share, then, understandably focuses mainly on the Conferences. Conference group interviewees discussed earlier in this chapter indicated that they appreciated the conferences, and the non-Conference group interviewees indicated that they felt they would have benefitted from having the teacher explain things directly to them, and indicated that they sometimes asked clarification questions. Thus, from the interviews with the student participants, there was no evidence that anxiety was the reason for their apparent reticence. However, it is possible that the student interviewees felt uncomfortable revealing their anxiety to me, as I was a stranger to them. Thus, I wondered whether the teachers had sensed that students felt anxious about their one-to-one conferences; interestingly, the answer was negative.

> Teacher 1 A couple seemed confused at first, about why [they were being given a conference], but not terrified. Mostly they just got on with it.

> Teacher 2 I think the students like the conferences. They seemed to be at ease. I don't know if they had expectations about what those conferences would be. I think they liked the conference.

Having seen the students' lack of participation in the conferences, it is remarkable that neither of the teachers indicated that students seemed uncomfortable; Teacher 2 even describes his conference attendees as seeming to be *at ease*. As the students participated so minimally in these conferences, it is intriguing that neither the students nor their teachers gave indication that the students were anxious about the conferences. This will be further explored in Chapter 7.

The two teacher participants differed in their ideas about, if asked to choose, which of the two interventions they would most likely implement in their own classes. Teacher 1's reason for preferring the worksheets to the conferences seemed to be a practical one – time.

> Teacher 1 You can't do a conference with each student, because you're going to have 19 very bored students waiting. You'd have to do it outside of class, really, if you wanted to have it with each student. Or you could plan another activity, I suppose. I prefer the worksheet just for that reason. … But it does take a long time to do the conferences. But the worksheets are all happening at the same time, so it's much easier to manage in each class. (In a) small, cozy class of 5 or something, it might not be an issue, but in a class of 20….

This is in line with other research, which reports that teachers agree that a disadvantage of conferences is the time-consuming nature of conducting one-to-one conferences with each of their students (Hyland, 2013a).

Teacher 2 explained his view that both of the interventions were superior to direct feedback. He seemed to indicate that in his experience, students would not examine their feedback unless forced to do so.

> Teacher 2 I mean if they're writing it out, and they're kind of recognising it in the end. I think that's an advantage, but I think it's good to talk to them. But both (conferences and worksheets) are better than just putting the corrections on paper. They won't look at it.

This view finds some support in the literature: of 121 final-year undergraduates, seemingly in the UK, majoring in primary education surveyed in Murtagh and Baker (2009), over a fifth of them (22%) reported rarely looking at the feedback they had received on their assignments throughout the programme. This is not always the case, though: in Carless (2006), interviewees with Cantonese L1 students (n=15) in a Hong Kong university and (like those in Murtagh and Baker) majoring in education (training to be English L2 teachers) reported that they did read and try to understand their feedback. This view seems more common in research, with multiple studies and reviews reporting that both L1 content course students (Zimbardi et al., 2017) and English L2 learners (Hyland, 2013b) value and want to understand feedback.

Despite acknowledging the possible benefits of using reflective worksheets, Teacher 2 preferred conferences. The main reason for this preference seemed to be the personalisation he felt it offered. He seemed to feel that the worksheets were just another handout, and that without engaging with the teacher, students could not maximise their learning.

> Teacher 2 I prefer conferences because you have the opportunity to have them ask questions and see their body language. You get a better feel for what they understand, for what they're doing. It's better for rapport in the classroom, and the better that is, the more people will ask questions. So I prefer conferences.

That the teacher considered the opportunity to ask questions to be a key benefit of the conferences is striking, given the overwhelming absence of questions asked, as seen in Chapter 5. Perhaps it was the opportunity itself and what the act of providing it may have communicated to the students that he felt was important rather than any questions that may be asked. The teacher mentions body language and rapport; the audio recordings were unable to indicate body language or give a full picture of rapport that might have been communicated between teacher and student during their

one-to-one interaction. It is possible that he felt that the cumulative build-up of rapport over the term would result in more questions asked. The conferences took place in Weeks 8 and 9 of the term; while it would seem this would have provided time for rapport to develop, it is possible that due to the foreignness of the teacher (Section 1.2.2), this amount of time was insufficient.

As Teacher 2 continues to explain his views on grammar teaching, he mentions the importance of providing students with frequent encounters of the grammatical forms.

> Teacher 2 I think the more that they have time to look at what they've done, the more grammar-related stuff they have, the better. The more frequency. The disadvantage [with worksheets] is that someone is just given a sheet, and you don't talk to them, then it's a disadvantage.

The teacher's views here echo some of the factors Schmidt (1990) identified as impacting on noticing (Section 2.3.1.3), such as 1) the frequency of encounters, which might be 2) made more perceptually salient through engagement with the teacher, 3) thereby increasing the learner's skill level.

While Teacher 2 seemed to agree with Teacher 1 about conferences being time-consuming, he pointed out that worksheets could also be quite time-consuming. He explained that if he were to implement the worksheet intervention in his classroom, he would feel obligated to mark the worksheets in order to ensure that students correctly understood the grammatical forms being targeted.

> Teacher 2 If I were to incorporate the worksheets, it would be my responsibility to (check the worksheets to ensure students were understanding the errors correctly). If the student found their error and they're writing the wrong correction, and they're doing this repetitively, then they'll keep doing that mistake until someone corrects it for them. And they'll have to unlearn what they've learnt. If you're going to use the worksheets, then it would be your responsibility to do that. So, yeah. Would I *like* the extra work? [laughs] Well, but it's (the teacher's) responsibility.

Conferences are time consuming, but as Teacher 2 insightfully indicates, Reflective worksheets can also be time consuming. It is logical that if the facilitating of feedback is to be a time-consuming endeavour, then it is essential to ensure that the feedback and the facilitation method used are both a) appropriate for the learners, and b) effective. It is possible that for feedback to be useful to learners in test-focused environments where there has been little writing practice even in L1 (as well as L2), then time must be spent helping learners adjust to receiving feedback and learn how to use it. Tian and Lowe's (2013) Chinese L1 graduate participants in a US university required a full term to

adjust to receiving and using feedback (Section 2.4.3.1), so it could be understood if, even with opportunities for one-to-one engagement with the teacher and guidance reflecting on feedback via reflective worksheets, the feedback provision and time to adjust to it in the current study was insufficient for the participants.

Teacher 1's answer to the question about students in the Reflective worksheet and Comparison groups asking questions about their feedback offers possible clues about why the grammatical accuracy of the three groups improved equally; the teacher's comment here supports statements made by students about non-Conference group participants seeking help (Sections 6.2.1.2 and 6.2.1.3).

> Teacher 1 (After class) I had quite a few questions when we gave the grammar feedback to students. You know, (like) at the conferences, some students didn't understand … (So) up to 5 minutes of people asking me questions. Some classes, no, but with others, quite a few people had quite a few questions.

Puzzlingly, it is unclear why students who were given explicit opportunity to interact one-on-one with the teacher generally seemed to resist doing so, while those who were encouraged to work alone sought out the teacher to ask questions and get additional explanations. It is possible that this was the result of the Reflective worksheet group having time to consider the issues being raised by the feedback, while the Conference group were 'live', feeling pressure to respond right away (*cf*. Section 2.2). It also seems possible that, having observed their peers having one-to-one conferences with the teacher, and having had time to think about their issues, they might have felt emboldened to ask questions.

### 6.3.2 Past feedback methods

The teachers came into the project having experimented with different methods of CF. Both teachers had mainly restricted themselves to teaching the grammar lessons that were in the textbook, and both had previously used forms of indirect feedback in their classes, but they had different approaches to providing that feedback. Teacher 1 used codes, and mentioned in a separate conversation that he usually corrected students' work at home, returning it in class.

Teacher 1 Error codes. So I would have like *a* for article, *p* for preposition, and they would have to go through and correct it themselves. And if they still weren't sure, they'd talk to me. [… ] If the student had loads and loads of errors, and they might be disheartened if I marked all of them, then I would just focus on things from class, but if they were doing really well, then I might… for some students, even if I marked every error on their papers, then they might only have 3-4 marks on the whole paper. For the better students it was less focused; for the lower students, it was more focused.

Perhaps with the idea of protecting students' self-efficacy, he chose to provide more or less focused feedback based on individual student performance, rather than applying either focused or unfocused feedback to all students. He seemed to feel that students would ask for help if they needed it, though given the Conference transcript analyses, this assumption might require review. Section 2.4.4.2 presented research indicating the importance of rewriting work (revisions) after receiving feedback, but Teacher 1 had not required students to produce revisions prior to this study.

Teacher 2, on the other hand, preferred to use class time to do both the writing and the marking of practice work. Like Teacher 1, he also had not required students to write revisions prior to this study; students were asked only to correct errors he identified on their work, i.e. indirect CF. If students were unable to correct the errors, Teacher 2 explained that his practice was to provide the answer and an explanation, i.e. direct CF supported by explicit statement of the grammar rule.

Teacher 2 Uhm, we used the grammar that was in the text. My method of correcting the grammar was, if I see something that's wrong, well, usually what I do is, we'll have some writing in class, and I, when they're finished writing, I'll have them come up to my podium and I will circle what's wrong and I will ask them to figure it out for themselves. And then when they come back the second time, what they've done correctly, I'll check it, and what they haven't done correctly, I'll give them the answer and explain what they've done wrong. And we do this several times throughout the semester.

The process he describes has some similarities with the conference protocol that was (unsuccessfully) applied in the current study, making the lack of FoI even more surprising. This suggests another reason for the lack of FoI seen in the current study: an unconscious mismatch between CF intent and CF implementation, which has been found in other studies as well (e.g. Ferris, 2006).

Teacher 2 seemed to feel that it was important to give students the opportunity to ask questions about their feedback. He acknowledged that one reason he marked student work one-to-one in class was that the students he taught tended to exhibit what he referred to as shyness, which he felt could preclude a student from seeking help if they wished unless time were set aside specifically for asking questions.

> Teacher 2 Just marking is not very good, not enough. Because one, they don't get to pose any questions. And two, if they have questions, then they're too shy to… to ask.

This observation is notable for two main reasons. Firstly, it suggests that the teacher might interpret the student silence seen in the conferences as a result of them being *shy*; interpretations of student silence will be discussed in the next chapter. Secondly, it suggests that the silence from students observed during the conferences is representative of what this teacher sees in the larger population. Neither of the teachers gave any indication that they found the silence seen in the conferences in the current study to be surprising or unusual.

## 6.4 Chapter summary

Having seen in the previous chapters that participants in all three groups in the study performed similarly, and that there might have been some unexpected factors which contributed to this, this chapter examined the experiences of the student- and teacher participants. In this chapter, student-participants talked about their current experiences in these classrooms and with the respective forms of feedback; they also acknowledged the lack of personal investment in their learning, and explained reasons why, despite wanting to be good at English, they felt unable to force themselves to spend time studying it. The data revealed that the Conference group participants might not have fully understood the purpose of discussing their errors with the teacher. The data also indicate that participants' English-learning backgrounds might have left them ill-prepared to take in and use even facilitated feedback. In the discussion chapter, I will consider these themes, including possible links between the learning backgrounds of the students and their apparent reticence in the conferences.

Finally, the chapter offered a view of the teacher-participants in the study and their experiences with the interventions. The teachers seemed to feel that both of the forms of facilitated feedback were more useful to students than the direct written CF received by the Comparison group. Encouragingly, they seemed to indicate their view that these interventions were feasible for real-world classrooms.

# 7 Discussion and Conclusions

## 7.1 Introduction

For the current study, a sample of Korean L1 STEM students in their first year of university (n=122) received instruction on select grammatical structures and then produced samples of new writing (considered as Pre-test, Post-test, and Delayed post-test) in academic genre over the course of one 15-week university term. Language analytic ability (LAA) and self-efficacy for understanding and using feedback were measured in order to see if these would predict improvement in grammatical accuracy; self-efficacy was measured again at the end of the study in order to see whether any of the interventions had helped participants increase their self-efficacy. Between the pre-test and the post-test, participants received two iterations of the intervention, facilitating their use of feedback of those select grammatical structures according to their intervention group: the Conferencing group received one-to-one conferences, where teachers were to scaffold participants in demonstrating an understanding about their grammatical errors; the Reflective worksheet group were asked to self-scaffold, using a reflective worksheet and the direct CF provided on their work; and the Comparison group were provided with direct WCF, but were not prompted to engage with their corrections or otherwise scaffolded in their understanding.

The results of this study reveal that Korean STEM English L2 first-year university students do not respond as expected to the facilitation of their use of corrective feedback, neither in terms of grammatical accuracy nor in terms of self-efficacy development; and that neither aptitude nor self-efficacy seem to predict grammatical accuracy outcomes. The purpose of this chapter is to examine the evidence for this by critically reviewing the findings, thereby extending the current knowledge of CF provision on L2 writing. Having summarised the study design and the main findings, the next sections will provide discussion of how the results of the study relate to the theory and literature, examining convergences and divergences with previous work in the field, and will consider possible reasons for these convergences and divergences, with theoretical and pedagogical implications of the findings being discussed. Following this, Section 7.4 will discuss the strengths and limitations of the study, and some possible directions for future research will be proposed in Section 7.5.

## 7.2 Discussion of research questions

The next sections discuss interpretations of the results reported in Chapters 4, 5, and 6. The discussion is organised by research question, starting with the RQs investigating differential effects of intervention group on outcomes (RQs 1-2), and then moving on to the RQs relating to mediating

variables hypothesised to impact on outcomes (RQs 3-4). An interpretation of the post-hoc research questions follows in Section 7.2.3. Interview data were focused on separately in Chapter 6; here the interview data are used throughout this chapter for triangulation and to illustrate the discussion.

### 7.2.1 Differential effects of facilitating feedback use on outcomes (RQs 1-2)

#### *7.2.1.1 RQ1 Effect of facilitating feedback use*

The treatments in the current study were selected because the literature suggests that a) indirect feedback may cause frustration in some participants, particularly for those of lower proficiency (Section 2.4.2.2); and b) while direct feedback may reduce this frustration by telling participants both precisely where the problems were and how to correct them, direct CF may fail to prompt the deeper processing of the feedback for learning (Section 2.3.1.2), meaning that CF with metalinguistic information could be a useful method. It was hypothesised that the two treatment groups, who were prompted to produce metalinguistic reflections and whose feedback use was facilitated, would be more likely to improve the accuracy of their use of the targeted structures in new writing than the Comparison group, who received direct CF only. This is predicted by Schmidt's (1990; 1995) noticing hypothesis, i.e. that more noticing can result in more learning; and by information processing Section 2.3.1.2), i.e. that attending to and engaging in deeper processing of the information (CF in this case) can encourage learning by prompting participants to reflect on the CF and the relevant rules. The interviewees' reports seem to suggest that these hypotheses are supported in the current study. It was seen in Section 6.2.1 that interviewees from each group seemed to be drawing on their explicit knowledge as they considered their CF (e.g. HBH2: 'I … see [the CF] and remember this and Ah! Ah!') and as they produced new writing (e.g. JSY1 and JJW3 reported focusing on remembering the CF, and JSL2 was able to identify a specific instance of CF helping her on a later writing assignment).

It was seen in Section 4.2 that the feedback delivery mode (facilitated or not) had no significant differential effect on participants' overall grammatical errors or article errors: all three groups improved significantly from pre-test to post-test, but differences between groups were not significant (RQs 1a and 1b). While this result was not the expected outcome, it cannot be taken to mean that the experience was not useful to the participants. As Ortega (2012:408) notes, 'any SLA notion of progress has always been non-linear, gradual, unevenly paced, and often proceeding through interim nontargetlike (but developmentally helpful) solutions'. Nonetheless, theory and research can be examined for possible explanations for the observed outcome. Section 3.6.6 explained the theoretical differences between the treatment groups and the Comparison group. All

groups were provided with opportunities to develop explicit knowledge about the targeted grammatical structures through instruction, and to practice using these structures in writing. All groups received CF designed to prompt noticing (Schmidt, 1990; 1995). Santos, López-Serrano, and Manchón (2010) note that the lack of significant difference between groups in their study might be explained by the lack of time pressure, i.e. the time given to participants might have been enough to 'have neutralized potential differences in noticing behaviour in the two treatment conditions'. Similarly, in the current study, it is possible that a) instruction, b) noticing-prompting CF, and c) required revisions together formed a sufficient condition for the development of explicit knowledge, and that further facilitation of CF use, at least in this population, may not add further benefit.

This study is not alone in finding a lack of difference between groups receiving different forms of CF. As discussed in Section 2.4.2.2, Bitchener and Knoch's (2008; 2009a; 2009b; 2010) treatment groups, (which include indirect CF; direct CF; oral/ written metalinguistic CF; or combinations of these) did not significantly differ from each other even though there was only one structure being targeted for feedback; each study also had a true control, which did not receive any CF on the targeted structures and whose results were significantly different from those of the treatments, in that the control demonstrated lower levels of improvement. In three of the studies (2008; 2009a; 2009b), one of the treatment groups was given direct CF; the authors suggest the possibility that for low proficiency learners, the form of CF is not as important as long as there is some provision of CF. Even with advanced-level ESL university students (mostly East and South Asian), there was no difference between treatment groups (Bitchener and Knoch, 2010).

However, not all CF studies fail to demonstrate a difference between treatment groups. As discussed in Section 2.4.2.2, Sheen (2007) found a significant effect for treatment over control for both direct CF and metalinguistic CF groups at the time of immediate post-test, and for metalinguistic CF at the time of the delayed post-test. Shintani, Ellis, and Suzuki (2014) found a significant improvement for one of their targeted grammatical features (hypothetical conditional) but not the other (indefinite article), and only at the time of the delayed post-test. As a possible cause, Shintani and colleagues suggest that asking the learners to focus on two structures – instead of one – might have placed high demands on learners' information processing resources, resulting in learners' inability to attend to the multiple structures. If that is indeed the case, the eight structures targeted in the current study, while likely a common number for focus in the classroom, might have had a similar effect. Shintani et al.'s study was published after the data collection of the current study, meaning that it could not

be used to inform the design; regardless, the current study strove for ecological validity, and teachers cannot spend an entire term on one or two grammatical forms.

It is possible that the results of the studies above finding no significant effect were affected by participant background. (Shintani et al. do report significant effects, but their dictogloss task is a very different type of task from the less structured writing of others in this discussion, and so this study is omitted from the following discussion.) The participants in all four studies by Bitchener and Knoch (no significant effect) were mostly East Asian (or, for the 2010 study, East and South Asian); participants in Sheen (significant effect) were comprised of three main L1 background groups, only one of which was an East Asian L1. If descriptions of prior English learning experiences given by interviewees in the current study were true for the East Asian participants in these studies (as might be the case (Williams, 2017)), then L1 background group might be indicative of participants' prior learning experiences in these studies, which might have influenced those participants' ability to use feedback. This issue will be explored further in Section 7.2.3.2.1.

The conversation thus far has taken the view the two treatment groups 'failed' to outperform the Comparison group, but since all groups improved their grammatical accuracy significantly from pre-test to post-test, it is possible to view the results from a different angle: it could be the case that the treatment groups' interventions were effective, and that there was something about the Comparison group that enabled them to perform *as highly as* the treatment groups. As noted in Section 6.3.1, it is possible that, observing some of their classmates having one-to-one discussions with the teacher, they felt emboldened to ask for help themselves. This help could have resulted in reducing any differences between groups. Another possible explanation for the 'unexpectedly high' (from this perspective) performance of the Comparison group is that overhearing grammar explanations in the conferences as they occurred in the classroom (Section 3.6.6.1) resulted in Comparison group members becoming attentive to the grammar features in question, and focusing on these independently. While a strength of the study design is that the groups are balanced across variables (Section 3.6.2), this meant that all three groups in the study were in the same classroom at the same time, which means Comparison group participants might have overheard the Conferences.

There are theoretical reasons, however, why this is an unlikely explanation. Information processing (Section 2.3.1.2) suggests that 'humans are limited-capacity processors' of information, and as such, must choose where to focus their limited attentional resources (McLaughlin, Rossman and McLeod, 1983:135). Students were busy with their own work (revisions), and teachers report that participants

did not exhibit visible signs of attention to the conferences or to their peers' work (e.g. straining to hear the conferences, reading over their neighbour's worksheets), so there are grounds for discounting intentional eavesdropping as a cause. Furthermore, participants' L2 proficiency (low B2) and the amount of time they required to produce each writing sample (about one hour to write 140 words (mean)) suggests the unlikelihood of eavesdropping as a cause. Nonetheless, however unlikely it might be, eavesdropping as an explanation for similar levels of improvement across groups cannot be entirely dismissed.

#### *7.2.1.2 RQ2 Effect of facilitating feedback use on self-efficacy development*

Self-efficacy is understood to have an effect on learning: if a student believes that they can succeed in a task, they may try harder to achieve the task than they might if they believed that success was unlikely (Bandura, 1977; Section 2.4.3.3). It was hypothesised that facilitating participants' use of CF on their errors might help participants develop their self-efficacy for understanding explanations of grammar and for using feedback, but this was shown not to be the case. A possible cause of this is context. In Section 2.4.3.3 it was explained that students need context in which to frame their estimates of their self-efficacy (Stajkovic and Sommer, 2000). It is possible, especially given the lack of L1 or L2 writing experience they reported having prior to university, that the writing completed during the current study was insufficient to provide sufficient context for estimating self-efficacy.

Other studies also report surprising findings, despite theoretical support in the literature for the connection between self-efficacy and learning. In Chen (2016), for example, the control group outperformed the treatment group in increased (pre- to post-test) writing self-efficacy. The small study involved Taiwanese undergraduates in English L2 courses divided into two groups: control (n=26) and treatment (n=18). Using university entrance exam results, the researcher estimated the participants' English L2 proficiency to be intermediate to high-intermediate. The participants in Chen (2016) were, like in the current study, mostly male and were STEM majors. The unexpected findings on self-efficacy in both Chen (2016) and in the current study indicate that there may be complex interactions between variables.

Some of these variables might include context. While researchers are increasingly acknowledging challenges of using and researching Western pedagogies in non-Western contexts (Casanave, 2009; Williams, 2017), similar concern about Western instruments commonly used in L2 research seems to have attracted less attention. There is reason to consider the issue. Some researchers propose that the Western construct of self-efficacy might not function the same in some Eastern contexts as in

some Western ones (Yu, 1996). Education is viewed very differently in, for example, the USA compared to how it is viewed in Korea. Section 1.2.1 highlighted some of those differences, such as possible Confucian influences on beliefs about education and the historical way in which educating the younger generation directly benefitted the older generation. Due to differences such as these, Kim and Park (2006) suggest that Western concepts of psychology can be less predictive in some Eastern contexts. As a result, Kim and Park emphasise the need to view psychological measurements (e.g. self-efficacy) and educational attainment within the cultural context. Research seems to support this. For example, Ho's (2009) examination of PISA exam and survey results for Hong Kong, Korea, Japan, and Macao found, though without statistical treatment, that Japanese and Korean students reported much lower self-efficacy for mathematics than the mean for OECD countries (and yet scored amongst the highest (OECD, 2010)); interestingly, this low self-efficacy result was not seen in Hong Kong or in Macao (also amongst the highest scoring), which the author posits may be a result of those countries being 'post-colonized Chinese societies, [which] might have absorbed some Western styles of parenting and schooling which may have had a positive influence on the tasks-specific self-perception of self-efficacy' (p. 338). This, too, suggests the need for caution when using and interpreting self-efficacy with Korean participants.

If currently available measures of self-efficacy are not the most appropriate for this population, *beliefs* might have been a more useful construct to use. Beliefs can 'broadly be defined as opinions and ideas that learners (and teachers) have about the task of learning a second/ foreign language' (Kalaja and Barcelos, 2006:1). Granted, there are some problems with this. Beliefs are considered to be a 'messy construct' (Pajares, 1992:307). This messiness prompted Dörnyei to omit *beliefs* from his 2005 book on individual differences in L2 learning, only adding it to the revised 2015 edition after he felt the construct had been more fully developed (Dörnyei and Ryan, 2015:188). Emphasising the dynamic nature of beliefs, Barcelos and Kalaja (2011) point out that beliefs, inter alia, a) fluctuate over time; b) are complex, and resist cause-and-effect analysis; c) can be influenced by others or by self-reflection; d) do not necessarily translate into action. As a result, '[t]here is no doubt … that our understanding of learner beliefs is far from complete (Dörnyei and Ryan, 2015:191). Due to the complicated relationship between belief and action, two considerations are 'to what extent are one's beliefs separate and separable from one's actions, and when can actions be said to be consistent with beliefs and under what circumstances can they be said to be inconsistent' (Woods, 2006:206).

The pre-intervention background survey results indicate that participants seemed to have beliefs about the usefulness of English L2 skills to their lives outside of the current classroom; participants identified a number of reasons why such skills would be useful, such as for their studies, for travelling, for getting a job – and some even indicated that they might need English for the work they might do for their job. However, despite evidence that they may have arrived at university with a desire to improve their English (Section 6.2.2), 87% of participants reported that they had spent zero hours each week studying English (*cf*. Section 6.2.2.2). Interviewees, despite being asked face-to-face, were forthcoming in their answers, replying with equanimity that they had not studied English. As discussed in Section 6.2.2, participants gave a number of reasons for this, such as not having the will to force themselves to study or giving in to peer pressure to do more enjoyable things with their classmates. Following Woods (2006), what appears to be a lack of effort could be characterised as an inconsistency between learner beliefs and learner action.

### 7.2.2 Mediating variables which might influence outcomes (RQs 3-4)

As discussed in Section 2.4.3, the literature suggests a possible role for aptitude and self-efficacy in language learning, so it was, at least initially, unexpected that neither LAA nor initial self-efficacy exhibited a significant mediating role in participants' improvement of either overall grammatical accuracy or accuracy in use of articles (RQs 3a and 3b). It is possible that there were contextual factors which affected the mediating power of LAA and self-efficacy. For example, the current study was conducted within the confines of the participants' normal classroom, and participants were 1) STEM majors, and 2) in required classes in their Korean university. The next sections will consider how these contextual factors may have impacted on the results of the tests.

#### *7.2.2.1 RQ3 Predictive value of aptitude and self-efficacy*

The potential for the Western construct of self-efficacy to function differently in non-Western contexts was discussed in Section 7.2.1.2; this problem could also be expected to be reflected here in RQ3, as the results were drawn from the same instrument. To avoid being repetitive, this section will focus on the construct of aptitude. Language aptitude was described in Section 2.4.3.4 as the 'ability to reflect on language and extract abstract information' (Ellis, 2004:251), and is considered to be a predictor of achievement (DeKeyser, 1993). However, this was not seen in the current study. As was discussed in Section 3.6.3.1.3, the LAA test in the current study was likely too difficult, so the results cannot give a reliable or well-differentiated picture of the underlying construct. This low predictive result can be considered a consequence of the choice of instrument (Section 4.2.4).

Putting aside the issues with the instrument in the current study, research suggests there are possible influencing factors which might be taken into consideration with future research using aptitude tests. For example, some groups might have a tendency to treat language more analytically. Rao and Liu (2011), discussed in Section 2.4.3.2, found that participants who were science majors indicated more study techniques relating to analysing language than social science majors; such tendencies might result in reducing the power of a language analytic test to differentiate between learners. Research suggests that the tendency to treat language analytically can be the result of English education: Shintani et al. (2014), though not focusing on aptitude, suggest of their Japanese L1 participants that 'as a result of many years of traditional grammar instruction, the learners in this study were oriented to treating English as a set of explicit rules' (p. 121). For reasons such as these, Kormos (2012:398) argues for the consideration of the 'social, cultural, and educational context' where L2 writing research takes place. She points out that affective factors, such as motivation, interests, attitudes, or self-beliefs are not stand-alone units created within classrooms, but rather that these can be influenced by the learners' linguistic, cultural, and social environment, and the experiences they bring with them.

There is some evidence that aptitude might be influenced by context as well. Indicating possible reason to question the 'stable trait' understanding of aptitude (Skehan, 1998; Section 2.4.3.4), Sáfár and Kormos (2008) found that, of their Hungarian secondary school-aged participants (n=54), those receiving intensive language instruction for one year significantly improved their aptitude scores from the beginning of the year, while their peers attending regular courses did not. The authors argue that the current understanding of aptitude needs to be re-evaluated in research in order to consider contextual factors. If contextual factors might influence aptitude, then future research would need to ensure that these factors were taken into consideration. One way of doing this might be to consider related factors. For example, research with Chinese L1 students suggests that these students tend to focus more on the importance of effort than of ability (Hau and Salili, 1991); Stevenson, Chen, and Lee (1993) found similar results for their Chinese L1 and Japanese L1 participants, but found that the American participants in their study demonstrated the opposite. Given four factors which could influence a test result (studying hard; good teacher; intelligence; home environment), only 27% of American participants selected *studying hard* as the most important factor, compared to 59% of Chinese students and 72% of Japanese students. Follow up questions to parents and teachers strengthen the case for a potential cultural link: given the same question, the Chinese and Japanese L1 teachers were also more likely than the American teachers to select *studying hard*, and 'American mothers, to a greater degree than Chinese and Japanese

mothers, emphasized the importance of innate ability' (ibid.:57). Clearly, within any culture, it would be expected that teachers, parents, and students would have a range of ideas about influencing factors of academic achievement. However, if one context demonstrates a tendency to favour effort over aptitude, and the other context demonstrates an opposite tendency, then it could be anticipated that the two contexts might be difficult to compare, something that might be important for future research to consider when attempting to make such comparisons.

This discussion is not intended to suggest that a test of aptitude might be replaced by measuring effort. Indeed, if the view is taken that, for this population, effort would be a better predictor than aptitude, then it would be curious that participants report having put forth so little effort. The argument here strives only to explore why aptitude might fail to predict improvement as expected even with properly calibrated instruments. It seems that constructs developed in the West might need to be carefully analysed in order to verify that they produce comparable results in the various contexts where the instruments are applied.

#### *7.2.2.2 RQ4 Effect of worksheet quality*

The current study applied two iterations of a process writing cycle: writing – receiving feedback – writing revisions – producing new writing. The two treatment groups received facilitation in using their feedback, during which they were prompted to reflect on the error and relevant rule. Information processing (Section 2.3.1.2) suggests that when learners' attentional focus is appropriately narrowed and when they are engaged in deeply analysing information, learning can be expected to be seen with time. According to the hypothesis, participants who produced higher quality reflections in the worksheets might be hypothesised as having engaged in deeper analysis, and thus might be expected to receive greater benefit from the activity. It is surprising, then, that for the Reflective worksheet group, worksheet quality (cf. Section 3.7.1.2) did not seem to influence participants' level of improvement in grammatical accuracy.

Though it is possible that it was simply a matter of two iterations being insufficient for differentiation between higher- and lower-quality metalinguistic reflections, it is also possible that the learners needed a deeper, more meaningful level of engagement with the task than they had. Theoretically, the task seems ideal for L2 development: the forms to be focused on have been made salient through instruction; and the range of focus has been limited for the learner, decreasing demand on their attentional resources; the participant is encouraged to notice errors, and to reflect on the underlying rules. This seems to correspond to the concept of deep processing, as it involves

more than just memorising (Section 2.3.1.2). An examination of the Worksheet scores, as reported in Section 4.2.6, suggests that they were not too difficult; and reports from the Reflective worksheet group interviewees in Section 6.2.1.2 suggest it is unlikely that the task involved 'too-deep' processing, as in Baralt (2013; *cf.* Section 2.3.1.2). Nonetheless, the mean worksheet quality score was 72.8% (Section 4.2.6); it cannot be ruled out that the task design might have been insufficient to prevent learners from filling out the worksheet with only surface-level engagement. Shintani and Ellis (2015), discussed in Section 2.4.2.2, reflecting on their results, suggest that when students write revisions after receiving either direct CF or metalinguistic CF, they might simply copy corrections without critical thought about what they are doing and why; this could fail to elicit the deeper processing hypothesised to promote learning, and thus might be an insufficient condition for L2 development. If participants in Shintani and Ellis can copy corrections without cognitively engaging with them, it is reasonable to think that the participants in the current study could copy information needed for metalinguistic reflections from their notes with similar low levels of cognitive engagement. This will be further discussed in Section 7.2.3.

### 7.2.3 RQs 6-9 Post-hoc research questions

As explained in Section 3.6.3.2.4, once the Conference transcripts had been analysed, it was clear that there were problems with fidelity of implementation (FoI), which resulted in post-hoc research questions being added to the study. The post-hoc RQs explored the implementation of the conferences and how students participated in the facilitated conditions. Two main sources were identified in Chapter 5 as contributing to the lack of FoI: one was the teachers' implementation of the conferences, and the other was the minimal or no response given by the students to the teachers' prompts. However, the quantification of errors and corrections is very much an SLA-influenced methodology. If examined from an L2 writing perspective (*cf*. Section 1.3), this attempt to impose a strict, quantifiable protocol for correcting errors has the effect of erasing, or attempting to erase, personal influences and interpersonal dynamics from the study.

> Once the analyst removes the interpersonal and social logic of a teacher's local decision to correct, the data can quickly turn into a disappointing record of teachers correcting in one way and then another way, correcting one student but not the next, correcting an error now but not correcting the same error later.
>
> (Ortega, 2012:412)

While the controlling of variables may be desirable for some purposes, it can be seen that doing so can also create a deficient view of the classroom, disregarding the complex synergies of interpersonal relationships and their potential influences on learning. Nonetheless, as this aspect of the study was designed with control in mind, factors relating to FoI stemming from the teachers and the students are considered in turn below.

#### *7.2.3.1 FoI: Teachers' perspective*

Though asked to follow a two-step protocol for conducting the conferences, an unexpected finding in the current study was that teachers failed to implement this intervention as it had been designed. This issue has already been identified as possibly contributing to the lack of statistical differences between the treatment groups and the Comparison group (Section 4.2.1). Oliver (1995) hypothesised that low levels of uptake of recasts in her study were often caused by either the interlocutor moving on before the learner had responded, or by the closed (yes/ no) nature of the questions being asked. If the lack of engagement can be explained as above, it could be said that the conferences had failed to engage the learner and prompt the intended deeper processing.

The exact reason underlying this lack of FoI cannot be known for certain; in Section 5.1, the familiarity of the term *Conferences* was indicated as contributing to the lack of FoI. A review of the literature suggests that another explanation may be that teachers' established pedagogical routines can prevent them from implementing pedagogical designs as intended. Types of behaviours 'where actions are repeatedly being performed without deliberating too much' are *habits* (Jager, 2003:149). As these teachers had been teaching and interacting with students in this context for a minimum of five years (Section 3.6.1.2), it might be said their established pedagogical routines had become habits. A consequence of habit formation is that, even when new information is provided, it is often disregarded; thus, habits tend to be resistant to change (Jager, 2003). For this reason, 'it is often very hard to change habitual information using a persuasive message' (ibid.:155). Habits can cause people to overlook new information because a result of the habit is that 'one may not recognise [the new information] as relevant for the [sic] own situation' (ibid.:155).

The advantage of habitualised routines is that they conserve cognitive processes through automatisation (Rushmer and Davies, 2004). Such automatisation can allow, for example, teachers to conduct conferences *while* being aware of the time left in class *while* noting how long each conference was lasting *while* gauging learners' levels of understanding and scaffolding them *while* keeping an eye on all of the students working on their revisions. The disadvantage of such routines is that, due in part to the same automatisation described above, they make *un*learning difficult (ibid.). If people are entrenched in their habitualised routines, they may 'dismiss (or simply fail to notice) information that conflicts with their view of the world' (ibid.:ii12). Having argued that unlearning requires explicit, purposeful investment of effort, Rushmer and Davies (2004:ii14) conclude that '[w]hat we already know (explicitly and tacitly) and what we are trying to assimilate, interact, and cross contaminate in non-linear and unpredictable ways'.

Given the level of teaching experience of the teachers in the current study, the voluntary nature of their participation, and the care which they exhibited in other aspects of the study, habitualised routines, as described above, rather than inability or unwillingness, seems the most plausible explanation for the disparity between what they were asked to do and what they actually did; the new information (conference protocol) was not able to overcome the existing information about what conferences were and how to conduct them.

Having examined how the teachers might have contributed to the problems of FoI, the next section looks at how student responses to the teachers might have caused the teachers to abandon the protocol adhering to it.

#### *7.2.3.2 FoI: Students' perspective*

A review of the data presented in Chapter 6, along with information about the context (Section 1.2) and the participants (Section 3.6.2) suggests that students' prior learning experiences may have contributed to the difficulty participants seemed to have engaging in the conferences. The background surveys and interviews revealed that students' prior experiences with English education were vastly different from what they were experiencing in the classroom in the current study. These possible effects of prior experience on conferences will be discussed in Section 7.2.3.2.1. Although prior experiences might also explain the participants' minimal answers or silences with the teacher seen in Chapter 5, research indicates that there could be other explanations for their silence, meaning it is worthwhile to examine this issue separately. For this reason, possible causes of silence will be explored separately from the discussion on prior experiences, in Section 7.2.3.2.2.

##### 7.2.3.2.1 Prior experience

The Introduction chapter provided information about the learning background that might be expected of the sample in this study, such as the test-focused tradition of English language instruction in schools. If students' educational and social contexts do not place high value on writing, then learners might approach the task of writing with a very different conceptualisation and valuation of it than learners whose writing activities had been emphasised (Cumming and Riazi, 2000). In the current study, for example, participants reported that L1 writing was not a large part of their secondary education, and interviewees described English L2 classes in secondary school in terms of *problem solving* and *selecting the right answers* on multiple-choice tests.

Research suggests that students moving from one educational system to another educational system might understandably struggle due to differences in priorities, values, and pedagogies (Blanton, 2005). Comments by interviewees in in Section 6.2.2 suggested that their prior English learning experiences were very different from their current English learning experience. For example,

> JSY1 'I.. frankly, I never use this writing…. Only right thing and wrong thing and select it'.

Though the data cannot firmly support conclusions drawn on this, it is possible to speculate that the learners' multiple-choice test-focused English education histories have primed them to engage best with multiple choices rather than with open-ended questions, as indicated by JSY1's comment about choosing the right answer. It was seen that learners who seemed unable or unwilling to answer open-ended questions asked by their tutors exhibited willingness to answer when the question was restated as a multiple choice question (e.g. *Is this noun a specific thing or a general thing?*). This could also be an effect of the teacher limiting the field, allowing the learner to focus their limited attentional resources on a narrower range of possible answers, as information processing might suggest as helpful (Section 2.3.1.2).

Some participants specified that they had been strategic in focusing their attention. For example, ChHJ3 explained that in secondary school, she had decided that she needed to be good at *some*thing when she entered university, and had decided English would need to be deprioritised.

> ChHJ3 제가 영어를 잘 봇하다 보니까 대학에 들어올려면 잘하는게 있어야 되잖아요. 그래서 영어 아예 포기하고… 다른 걸 열심히 했었고. 영어를 잘못해서…. / I realised I wasn't good at English, but when you go to university, you have to be good at something. So I completely gave up on English and worked hard at other things, since I was bad at English.

This participant is unique because her self-efficacy scores are the lowest of the interviewees: this participant's estimated her initial self-efficacy at 32%. As reported in Section 4.2.3, the mean initial and final measurements for self-efficacy for the Conference group were 73.4 and 73.8 respectively, so this participant's low self-efficacy scores are strikingly low in contrast. However, she is not alone; approximately 10% (n=11) of 106 self-efficacy scores were under 50%.

Examining ChHJ3's background survey provides some clues. It can be seen that she began studying English around the age of 5, and that in addition to having English classes at school from kindergarten through the final year of secondary school, she also estimated that they attended English cram schools for five years and had private tutoring in English. With such intensive English

instruction, ChHJ3's L2 proficiency, reluctance to use the L2, and low self-efficacy might seem surprising, but a deeper look reveals problems that can happen when the drive to learn English becomes too high:

> [There can be a] fierce competitiveness among parents to push their children's English acquisition quickly, perhaps to impress their friends by displaying their child's highly advanced workbook. Many teachers report pressure from parents to promote children to higher-level classes […] but then the children feel lost and frustrated in a class that's too hard for them.
>
> (Bizzell, 2017:76)

Williams (2017:107) discusses the need for teachers to use Western pedagogies with learners' cultural and educational backgrounds in mind. The conference intervention was structured to accommodate learners' discomfort with communicating in English L2, but this may have been an insufficient adaptation given the participants' prior learning experiences. Firstly, they had surprisingly little L1 writing experience, which would presumably translate into equally little experience receiving feedback. Secondly, as reported in Section 6.2.2, several interviewees reported that they had never received feedback on L2 writing, and that their classes had been test focused – understandable, given the importance of the CSAT (Section 1.2.2). Research suggests that other populations in, e.g. the East Asia region have similar low levels of L1 writing experience. For example, Reichelt (2009:196) reports that Japanese secondary school students 'do not perceive EFL writing as important for their future study or careers', and that non-English majors in Japanese universities 'often do not perceive a need for writing in English'. Though there is evidence that this is changing in some places in Japan (Rinnert and Kobayashi, 2009), it seems consistent both with the current research and with other reports in the region (Williams, 2017).

Though it was not feasible to interview participants' secondary school teachers for evidence of their test-focused learning background, there is some support for the participants' assertions about this in the literature. Guilloteaux (2013) administered two Likert-style surveys gathering teachers' opinions about pedagogical strategies to Korean L1 secondary school teachers of English L2 (n=268); half of the participants received the survey asking about the *importance* of items (Not at all important – Very important) and the other half received a survey asking about the *frequency* with which teachers performed the item (Hardly ever – Very often). Results indicate that teachers ranked survey items relating to promoting learner autonomy the lowest for both measures. The teachers rated the *importance* of facilitating (as opposed to lecturing) at 4.9/6.0, but the *frequency* of facilitating in the classroom was only 3.1/6.0. The author explains that '[t]his result corroborates findings that, while

Korean teachers may acknowledge the need to switch to a more communicative approach, they continue to mostly lecture about the language rather than teach how to use it' (Guilloteaux, 2013:12). It is worth noting that these surveys were administered in the context of four required workshops being delivered by the researcher, two of which counted toward promotions for participants, and so it is reasonable to think that some of the teachers might have inflated their answers slightly. This notion is further strengthened by the 100% return rate of these optional, anonymous surveys.

If teachers who might feel internal pressure to skew their responses toward the 'right' answers, as above, report only *Sometimes* encouraging students to ask questions and only *Sometimes* facilitating rather than lecturing, then this, combined with the indications from the participants in the current study that they had little L1 or L2 writing (and thus, feedback) experience, seems to indicate that learners in this population might be ill-equipped to interact with their teachers about their work. Teachers wishing to implement conferences with this population might find it necessary to scaffold students to enable them to reach this level of interaction. If participants in the current study have experienced only classrooms where teachers lectured and students took notes, then they might have struggled to understand the value of the conferences or of focusing on their errors. In a test-centric system, perhaps they found this 'post-mortem' focus on errors to be too little, too late, and failed to see how they could use the feedback to improve later work.

There are numerous ways prior experience could influence learners' writing and uptake of CF. In a review of literature Wigglesworth and Storch (2012) suggest that learners' attitudes toward accuracy could impact on how they engage with CF. Another problem stemming from the participants' prior experience is their difficulty judging their own accuracy in English, as witnessed by their expression of surprise that their work contained errors (*cf.* Section 6.2.2.1). It might seem puzzling that students in the process of learning a language would be so surprised upon the discovery that their work contained errors. To understand this, it may be useful again to consider context, and its impact on learners' notions of what learning an L2 entails. Depending on where they live, English L1 speakers might be accustomed to hearing their L1 being used by L2 speakers – for example, in their schools, in society, and on TV. English L1 speakers might develop a tacit understanding though encounters with non-standard expressions – tenses, articles, etc. that when one learns an L2, non-standard use will occur. In Korea, however, such encounters are not generally the case. As the Introduction chapter reported, Korea is a highly monolingual society; anecdotal evidence and the public service advertisements mentioned in Section 1.2.2 suggest that many Koreans go about their

days never seeing anyone who is noticeably foreign, so encountering someone using Korean L2 might be considered very rare. This means that Korean L1 learners of English might not have experience with the non-standard utterances language learners produce. This lack of awareness could make it difficult for them to realise that L2 speech tends to contain errors. This could have implications for using measures of self-efficacy with this population; if they do not possess an awareness of the errors that occur in L2 use, they might overestimate their ability to perform with accuracy in the L2. This concept of being unaware of what one does not know is referred to in the field of psychology as the *Dunning-Kruger effect* (Kruger and Dunning, 1999); suggestions for applying this concept to language learning research will be explored in Section 7.5.

#### 7.2.3.2.2 Silence

Having discussed participants' prior learning experiences and how these might have impacted on the study, this section will focus specifically on silence, looking to the literature to provide understanding. As was seen in Section 5.2.2, Conference group participants generally seemed reluctant to engage with the tutors; tutor questions were often met with minimal answers or with silence. Silence had been a problem in the conferences in the first pilot study; it was because of this that the conferences in the current study took place in a full classroom rather than in an empty one, as the empty room had seemed to intimidate the pilot participants. The pilot experience also contributed to the decision to use structured conferences rather than open ones, asking specific questions (*Why should we use ate instead of eat?*) as opposed to more general ones (*Do you understand?*). Despite these accommodations, most of the participants maintained their silence, leaving the tutor to interpret the meaning of the silences.

A range of things silence might indicate can be found in the literature. Focusing on the Korean context, Jang (2006) offers a discussion of Myers and Myers' (1992) identified meanings that silence might carry, which Bao classifies and labels. These are provided below in the first two columns of Table 7-1, with the third (shaded) column indicating whether there was evidence of each in the data of the current study; these are addressed in the discussion below.

**Table 7-1 Bao's (2014) classifications of silence and indication of likelihood of presence in the current study**

| Silence type | Examples of what it might indicate | Evidence in current data |
|---|---|---|
| Considerate | Politeness; respect | Unlikely |
| Positive | Agreement; understanding | Unclear |
| Insensible | Carelessness; distraction | Unclear |
| Introverted | Low communication skills; passivity | Likely |
| Negative | Disagreement; rebellion | Possible |
| Emotional | Anxiety | Likely |
| Uninformed | Not knowing the topic | Likely |
| Indifferent | Contempt; low interest | Possible |

The possibility of each of these being a cause of silence in the conferences in the current study will be considered, beginning with the ones that might be dismissed. Bao (2014:93) writes that in Korea, students tend not to initiate conversation with their teacher, but 'wait for signs of nomination'. In the one-to-one conferences, teachers were, it would seem, clearly nominating the student, so *Considerate* might be eliminated as a major cause of silence in this case. Whether or not any students exhibited *Positive* silence, indicating agreement or understanding, cannot be extrapolated from the existing data; likewise for *Insensible* silence. Examples given for the *Introverted* silence type indicate that this can relate to 'low communication skill'. In an L2 context, this might be interpreted as L2 proficiency. It would be understandable if low English L2 proficiency prevented the students from participating in the conferences, but given the intermediate-level L2 proficiencies seen in Section 3.6.2, it seems that proficiency alone is insufficient as an explanation for the frequent lack of participation in the conferences. Another interpretation of *Introverted* is possible. That this population can have difficulty interacting with non-Koreans was demonstrated in Section 1.2.2; if the Conference participants felt uneasy engaging in one-to-one interaction with their Western teacher, it is possible that the *Introverted* type of silence could have played a part in the participants' lack of interaction.

There were very few instances when a student showed surprise or uncertainty about a correction during the conference, but it is possible that some students kept this thought to themselves rather than verbalise it, so this could be classified as *Negative* silence, indicating disagreement. Recalling from Section 2.4.3.1 the experiences of Tian and Lowe's (2013) students, it seems that some of the silence could stem from anxiety, and thus be categorised as *Emotional*. It could be seen in the conferences that some students were willing to admit that they did not know the answer, so it is

reasonable to assume that others might also have not known the answer and remained silent (*Uninformed*). Of the 122 background survey respondents, 21 expressed a lack of enjoyment of studying English, so low interest could contribute to silence for some students (*Indifferent*).

This section has relied on literature on silence to understand why, apart from prior educational experiences, participants might have been unwilling or unable to engage more actively in the conferences. Particularly probable causes were feeling uneasy engaging in one-to-one interaction with the Western teacher, and not knowing the answers, in which case participants might have felt more comfortable waiting for the teacher to offer help than speaking up and requesting assistance.

## 7.3 Implications for research and pedagogy

A number of theoretical, research, and practical implications arise from this study. Truscott's (1996) claim that CF was harmful (Section 2.4) has not been supported in this study, as all groups significantly improved their grammatical accuracy. However, the results indicate the complexity of the relationship between CF provision and L2 (grammatical) development (*cf*. Section 2.3). Participants seem to report that providing feedback prompted their noticing (*cf*. Section 6.2.2.1), and the treatment groups seem to report drawing on explicit knowledge to deal with their CF. However, facilitating learners' use of the CF by prompting them to analyse the CF and process it did not seem to produce greater L2 development than direct CF alone. The methodological issues with the instruments and the problem of FoI make theoretical arguments difficult to support or refute. Nonetheless, there are implications both for the conduct of research in topics such as this, and with regards to substantive issues of CF use. To begin with the research standpoint, the results confirm the importance of a mixed methods approach to CF investigations. Without consideration of participants' backgrounds, conference transcripts, and interviews, the pre-test, post-test, and delayed post-test results would have provided only a small part of what was learnt in this study. The subtle nuances of all that surrounds the concept of learning cannot be captured solely by using test results.

Secondly, the study suggests that even experienced practitioners who are highly motivated to implement classroom interventions according to a clear protocol can struggle to do so, possibly as a result of a barrier to understanding formed by teachers' prior knowledge and experiences. The implications of this apply both to researchers and to materials developers. When explaining a classroom intervention, it might be useful to first investigate teachers' existing understanding of the proposed intervention and use that information to highlight any differences that might exist

between teachers' current understanding and the desired method of implementation. The usefulness of this recommendation can be seen for any intervention, but especially for interventions that are similar but not identical to teachers' established pedagogical routines; teachers' misperception of their understanding might be more likely for familiar-sounding interventions than when the intervention is perceived as novel.

In addition to issues of intervention implementation discussed above, the results of the study suggest that higher levels of engagement than those seen in the current study may be necessary for learners to benefit from scaffolded learning. For students like those in the current study, whose educational background has been overwhelmingly test-focused and whose experiences with teaching may not have included scaffolded learning, the ability to engage at those higher levels may be a skill that requires training. If there is not space for this training in the curriculum, L2 development may still occur, with instruction, direct CF, and required revisions being sufficient conditions for prompting enough noticing for some L2 development. Classroom teachers whose schedules do not allow time for CF-use training might welcome the finding that CF provision alone can statistically significantly improve learners' grammatical accuracy without requiring the extra work of also facilitating CF use.

## 7.4 Strengths and limitations of the study

There are a number of strengths in the design of this study which contribute to its ability to inform both theory and pedagogy, but also limitations which must be taken into account when interpreting the findings. Firstly, while much CF research has focused on types of feedback, the current study contributes to the small but growing body of knowledge on how CF can be delivered more usefully to learners. Secondly, to investigate this, the study used a mixed methods design with sufficient numbers of participants for statistical inferences to be made. The mixed methods design allows not only for measurements of the effectiveness of an intervention, but also for consideration of what student and teacher experiences with the intervention were. Thirdly, the study takes place in a foreign-language setting with STEM majors in required English L2 classes, and a) does not assume that participants are highly motivated to learn the language; b) does not assume that participants' L2 learning goals are to immerse themselves in an English L2 environment; and c) does not assume that participants perceive an immediate need for English L2 beyond obtaining a passing mark in the course. In this way, the results of the study contribute to the growing field of instructed second language acquisition (Long, 2017). Fourthly, the study was designed with ecological validity in mind: a) it investigates L2 development over the course of a term, as would be expected in a typical

classroom; b) it includes instruction on a range of grammatical structures and then focuses on the development of only those structures; c) all measurements are taken from new pieces of academic writing (as opposed to revisions); d) the classroom teachers implemented the research design, allowing it to be seen how my instructions were understood, adapted, and implemented in the classroom.

This ecological validity is a strength of the study; it was not designed to be a true experiment, but this impacts on the interpretations drawn. From a pedagogical standpoint, the lack of a no-treatment control group might be considered unproblematic, as few teachers would advocate withholding feedback from their students (Ferris, 2004). However, theoretical discussion is unfortunately somewhat limited by the lack of a true control. Another limitation that is the converse of one of the strengths above is the implementation of the design by the teachers; had I been able to implement the design myself, greater control would have been maintained over the implementation. Even though the current design allows insight into teachers' use of pedagogical interventions, the lack of control over the implementation of the interventions limits both theoretical discussion of the interventions and the kinds of pedagogical implications that can be drawn. While this lack of control over the implementation of the interventions represents a weakness, a corresponding strength is that the 'buffer' between the researcher and the student participants allows for more objectivity than might have been possible had I implemented my own design with my own students.

## 7.5 Future research

As is the case with much research, the results of this study raise a number of questions for further exploration. The results of this study also indicate that more work needs to focus on learner beliefs, particularly looking at the issue from a global perspective. For example, studies might identify the beliefs of populations similar to those in the sample of the current study and compare the findings to those found for other populations, both similar and dissimilar ones. In particular, it seems it would be useful to investigate the disparity between beliefs and action, and to identify causes of disparity and under what conditions it is most and least severe. It would be useful to investigate whether there are ways of minimising its effects. Such research might investigate the effectiveness of awareness-raising and training activities on reducing the beliefs—action gap. Relatedly, an issue suggested by the results of this study is the possibility that some western psychometric instruments might not be appropriate in some Eastern contexts. This echoes Goldstein's (2016) call for more research on contextual factors, such as how prior experiences affect students' use of feedback. This

is a critical point for research, as the results of these studies could have serious implications for studies being conducted with populations in Eastern cultures.

Section 7.2.3.2.1 suggested the possibility that awareness of what L2 language learner speech 'sounds like' might help learners be aware of errors in their own language production. The *Dunning-Kruger effect* (Kruger and Dunning, 1999) would be a useful avenue for research in language learning to explore, relating it, for example, to self-efficacy measures or to noticing. To offer an example, a study could investigate Korean L1 learners' attitudes toward or noticing in their English L2 before and after, for example, engaging them in an awareness-raising activity, such as working with Korean L2 learners for a period of time on a project to give them exposure to errors made by learners of languages.

Future teacher professional development research might investigate fidelity of implementation, seeing how teachers' established pedagogical routines can influence how they incorporate new pedagogical techniques into their classrooms and discovering what kind of training is necessary to disrupt these routines.

Whatever future research is conducted, the results of this study emphasise that longitudinal, mixed-methods research can allow researchers to obtain nuance and depth that would be lacking were the study to take only a quantitative or qualitative approach. As researchers have argued (Atkinson, 2002), and as this study suggests, the act of learning a language is at the same time social and cognitive. The current study suggests this includes instructed settings where learners might have only limited exposure to the target language outside of the classroom. Learners' context and prior experiences, particularly their educational experiences, shape their expectations for and actions in the classroom. Because of the influence of the social on the cognitive, and of the cognitive on the social, studies of language acquisition need to examine both in order to obtain the fullest picture possible.

## 7.6 Conclusion

The study reported in this thesis investigated how facilitating Korean STEM English L2 first-year university students' use of corrective feedback might differentially improve the accuracy of their use of targeted grammatical structures compared to a comparison group whose feedback use was not facilitated. The two treatment groups receiving facilitation either participated in one-to-one conferences or completed reflective worksheets; the Comparison group received direct WCF without

facilitation. As described in Section 3.6.6, the cycle described in Gass' (2018) interaction hypothesis and adapted by Bitchener and Storch (2016) for CF begins with receiving CF input (*noticing*) and comprehending the CF and taking it in. The two treatments were hypothesised to help students develop their explicit knowledge of the targeted grammatical forms by facilitating deeper processing through languaging (that is, by eliciting metalinguistic explanations about errors). To consider the interventions from a skills perspective, having established explicit knowledge 'declarative crutches' (DeKeyser, 1998:49), practice opportunities for writing were hypothesised to help proceduralise this explicit knowledge. Through this, participants might integrate the explicit knowledge, which might manifest in their new output. Thus, both from an interaction perspective, and from a skills perspective, the two treatment groups were hypothesised to demonstrate similar levels of improvement in grammatical accuracy of targeted structures over that of the Comparison group, indicating that reflective worksheets could be used as an effective alternative to time-consuming conferences.

Before beginning the intervention, participants were asked to complete a background survey and were administered pencil-and-paper tests to measure self-efficacy and language aptitude; these tests are suggested in the literature to mediate language learning. Prior to the intervention, eight structures were selected by a needs analysis of work completed by students in previous cohorts. Five of these were selected for instruction because it could not be guaranteed that they had been taught to learners before; the other three (basic-proficiency level) features could be presumed to have been taught in the students' school curriculum. All eight grammatical features were targeted for CF. To test whether the two ways of facilitating CF use enhanced learners' ability to improve their grammatical accuracy, participants produced three writing samples, considered as the pre-, post-, and delayed post-tests. Improvement in accuracy of targeted structures was measured using error ratios, with the two iterations of treatment delivered between the pre-test and the post-test. The results reveal that there was no advantage for any of the groups; while the two treatment groups demonstrate significant improvement in the accuracy of their use of the targeted grammatical features, unexpectedly, the Comparison group participants also made equally significant gains.

This thesis has extended the current knowledge of CF provision on L2 writing with the finding that Korean STEM English L2 first-year university students do not gain the additional benefit expected from having their use of CF facilitated. The findings highlight the complexity of feedback and the need for more empirical work to be done to identify the factors which affect the effectiveness of

feedback. There is need for theory that can account for these factors, including cultural and contextual factors.

This chapter discussed these results, highlighting a number of factors which might have contributed to this finding: a) the way the Conferences were conducted and participants' proficiency in or comfort using English L2 might have lessened the effect scaffolding was hypothesised to offer; b) participants' prior learning experiences, namely lack of writing and receiving feedback, might mean that they require training in receiving and using CF before the full benefit of CF can be experienced. The study also found that neither aptitude nor self-efficacy seem to predict the improvement in grammatical accuracy. The literature suggested that culture may play a mediating role in both self-efficacy and in language aptitude. Implications of these findings for both research and pedagogy were discussed, followed by a consideration of the limitations of the study. Possible avenues for future research were explored.

It is essential to emphasise that this study is not criticising the secondary school and academy teachers of these participants. The current reality in South Korea is that university rankings are very important for academic, professional, and social opportunities. To enter a highly ranked university, high CSAT scores are needed. To achieve high CSAT scores, students need to be able to do well on a multiple choice test. To help students do well on multiple choice tests, teachers must teach in a way that will help students master these tests; to do otherwise would be doing their students a great disservice.

This study was undertaken a) with the idea that students entering university need to be equipped to handle the different English learning context they are entering; and b) to contribute in a small way to identifying how best to reach these learners and help them to reach their personal language-learning goals. It would be erroneous for practitioners and researchers alike to assume that all learners are – or should be – striving toward the highest level of L2 proficiency, but if specific groups of learners are not responding as expected to an intervention, like CF, it may be that research needs to explore what it is about those groups that might be contributing to the observed results, and to consider which interventions and instruments are best suited to the group being researched.

# References


Abry, T., Rimm-Kaufman, S.E., Larsen, R.A. and Brewer, A.J. (2013) The influence of fidelity of implementation on teacher–student interaction quality in the context of a randomized controlled trial of the Responsive Classroom approach. *Journal of School Psychology,* 51 (4), pp. 437-453.

Afitska, O. (2015) Role of focus-on-form instruction, corrective feedback and uptake in second language classrooms: some insights from recent second language acquisition research. *The Language Learning Journal,* 43 (1), pp. 57-73.

Ahmed, M. (1994) Speaking as cognitive regulation: A Vygotskian perspective on dialogic communication. In *Vygotskian Approaches to Second Language Research*, eds. J. Lantolf and G. Appel, London: Ablex Publishing, pp. 157-171.

Akakura, M. (2012) Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research,* 16 (1), pp. 9-37.

Aljaafreh, A. and Lantolf, J.P. (1994) Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal,* 78 (4), pp. 465-483.

Altman, D. (1991, 1999) *Practical Statistics for Medical Research.* Reprint ed. Boca Raton: Chapman & Hall/CRC.

Alvesson, M. (2003) Beyond neopositivists, romantics, and localists: A reflexive approach to interviews in organizational research. *The Academy of Management Review,* 28 (1), pp. 13-33.

Anderson, J. (1982) Acquisition of cognitive skill. *Psychological Review,* 89 (4), pp. 369-406.

Anderson, J.R. (1983) *Architecture of Cognition.* Electronic 2009 reprinted ed. New York: Psychology Press.

Andrews, K. (2007) The effects of implicit and explicit instruction on simple and complex grammatical structures for adult English language learners. *TESL-EJ,* 11 (2), pp. 1-15.

Atkinson, D. (2002) Toward a sociocognitive approach to second language acquisition. *The Modern Language Journal,* 86 (4), pp. 525-545.

Attali, Y., Saldivia, L., Jackson, C., Schuppan, F. and Wanamaker, W. (2014) Estimating item difficulty with comparative judgments. *ETS Research Report Series,* 2014 (2), pp. 1-8.

Baek, S., Song, J. and Lee, J. (2012) Korean middle school students' cognitive processes during English writing. *English Teaching,* 67 (2), pp. 183-205.

Bandura, A. (1977) Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review,* 84 (2), pp. 191-215.

Bandura, A. (1984) Recycling misconceptions of perceived self-efficacy. *Cognitive Therapy and Research,* 8 (3), pp. 231-255.

Bandura, A. (1997) *Self-Efficacy: The Exercise of Control.* New York: W.H. Freeman and Company.

Bandura, A. (2006) Guide for constructing self-efficacy scales. In *Self-Efficacy Beliefs of Adolescents*, eds. F. Pajares and T. Urdan, Greenwich: Information Age Publishing, pp. 307-337.

Bange, P., Carol, R. and Griggs, P. (2005) *L'apprentissage d'une langue étrangère: Cognition et interaction.* Paris: L'Harmattan.

Bao, D. (2014) *Understanding Silence and Reticence: Ways of participating in second language acquisition.* Electronic ed. London: Bloomsbury Academic.

Baralt, M. (2013) The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition,* 35 (4), pp. 689-725.

Barcelos, A.M.F. and Kalaja, P. (2011) Introduction to Beliefs about SLA revisited. *System,* 39 (3), pp. 281-289.

Barker, M. and Pinard, M. (2014) Closing the feedback loop? Iterative feedback between tutor and student in coursework assessments. *Assessment & Evaluation in Higher Education,* 39 (8), pp. 899-915.

Bartimote-Aufflick, K., Bridgeman, A., Walker, R., Sharma, M. and Smith, L. (2016) The study, evaluation, and improvement of university student self-efficacy. *Studies in Higher Education,* 41 (11), pp. 1918-1942.

Beaumont, C., O'Doherty, M. and Shannon, L. (2011) Reconceptualising assessment feedback: a key to improving student learning? *Studies in Higher Education,* 36 (6), pp. 671-687.

Berliner, D. (1988) *The Development of Expertise in Pedagogy.* Washington, DC: AACTE Publications.

Best, K., Jones-Katz, L., Smolarek, B., Stolzenburg, M. and Williamson, D. (2015) Listening to our students: An exploratory practice study of ESL writing students' views of feedback. *TESOL Journal*, 6 (2). pp. 332–357.

Beuving, J. and de Vries, G. (2014) *Doing Qualitative Research.* Electronic ed. Amsterdam: Amsterdam University Press.

Bialystok, E. (1982) On the relationship between knowing and using linguistic forms. *Applied Linguistics,* 3 (3), pp. 181-206.

Bialystok, E. (1988) Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology,* 24 (4), pp. 560-567.

Bialystok, E. and Barac, R. (2012) Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition,* 122 (1), pp. 67-73.

Bickhard, M. (2005) Functional scaffolding and self-scaffolding. *New Ideas in Psychology,* 23 (3), pp. 166-173.

Bickhard, M. (1992a) How does the environment affect the person? In *Children's Development within Social Context, V.1: Metatheory and Theory.*, eds. L. Winegar and J. Valsiner, New Jersey: Lawrence Erlbaum Associates, pp. 63-92.

Bickhard, M. (1992b) Scaffolding and self-scaffolding: Central aspects of development. In *Children's Development within Social Context, V2: Research and Methodology*, eds. L. Winegar and J. Valsiner, New Jersey: Lawrence Erlbaum Associates, pp. 33-52.

Biesta, G. (2010) Pragmatism and the philosophical foundations of mixed methods research. In *SAGE Handbook of Mixed Methods in Social and Behavioral Research.*, eds. A. Tashakkori and C. Teddlie, 2nd ed. London: Sage Publications, pp. 95-117.

Bitchener, J. and Knoch, U. (2008) The value of written corrective feedback for migrant and international students. *Language Teaching Research,* 12 (3), pp. 409-431.

Bitchener, J. and Knoch, U. (2009a) The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics,* 31 (2), pp. 193-214.

Bitchener, J. and Knoch, U. (2009b) The relative effectiveness of different types of direct written corrective feedback. *System,* 37 (2), pp. 322-329.

Bitchener, J. and Knoch, U. (2010) Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing,* 19 (4), pp. 207-217.

Bitchener, J. (2016) To what extent has the published written CF research aided our understanding of its potential for L2 development? *International Journal of Applied Linguistics,* 167 (2), pp. 111-131.

Bitchener, J. and Storch, N. (2016) *Written Corrective Feedback for L2 Development.* Kindle ed. Bristol: Multilingual Matters.

Bizzell, P. (2017) Who owns English in South Korea? In *Crossing Divides: Exploring translingual writing pedagogies and programs.*, eds. B. Horner and L. Tetreault, Electronic ed. Boulder: Utah State University Press, pp. 70-86.

Björkman, B. (2008) 'So where we are?' Spoken lingua franca English at a technical university in Sweden. *English Today,* 24 (2), pp. 35-41.

Blanton, L. (2005) Student, interrupted: A tale of two would-be writers. *Journal of Second Language Writing,* 14 (2), pp. 105-121.

Blau, S. and Hall, J. (2002) Guilt-free tutoring: Rethinking how we tutor non-native-English-speaking students. *Writing Center Journal,* 23 (1), pp. 23-44.

Boggs, J. (2013) *The differential effects of corrective feedback delivered on paper compared to corrective feedback delivered live in conferences for EFL natural science students in required English L2 writing classes.* Unpublished dissertation ed. Oxford: University of Oxford.

Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3 (2), pp. 77-101.

Breakwell, G. (2012) Interviewing. In *Research Methods in Psychology*, eds. G. Breakwell, J. Smith, and D. Wright, Fourth ed. London: Sage, pp. 367-390.

Brinkmann, S. and Kvale, S. (2015) *InterViews: Learning the craft of qualitative research interviewing.* Third ed. London: Sage.

Brooke, K., Aden, M., Al-Kuwari, N., Christopher, V., Ibrahim, M., Johnson, B. and Souyah, O. (2012) Placement testing in an EFL context. *TESOL Arabia Perspectives,* [Online]. 19 (2), pp. 3 March 2013-13-20. Available from: <http://www.ucalgary.edu.qa/files/Article%20in%20Perspectives%20June%202012.pdf>.

Brown, H. (2007) *Principles of Language Learning and Teaching.* 5th ed. White Plains: Pearson.

Brown, D. (2016) The type and linguistic foci of oral corrective feedback in the L2 classroom: A meta-analysis. *Language Teaching Research,* 20 (4), pp. 436-458.

Bruce, E. and Hamp-Lyons, L. (2015) Opposing tensions of local and international standards for EAP writing programmes: Who are we assessing for? *Journal of English for Academic Purposes*, 18, pp. 64-77.

Bruning, R., Dempsey, M., Kauffman, D., McKim, C. and Zumbrunn, S. (2013) Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology,* 105 (1), pp. 25-38.

Bruton, A. (2009) Improving accuracy is not the only reason for writing, and even if it were…. *System,* 37 (4), pp. 600-613.

Bryman, A. (2012) *Social Research Methods.* 4th ed. Oxford: Oxford University Press.

Butler, R. (1988) Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology,* 58 (1), pp. 1-14.

Carless, D. (2006) Differing perceptions in the feedback process. *Studies in Higher Education,* 31 (2), pp. 219-233.

Carnicelli, T. (1980) The writing conference: A one-to-one conversation. In *Eight Approaches to Teaching Composition.*, eds. T. Donovan and B. McClelland, ERIC: http://eric.ed.gov/?id=ED191042: National Council of Teachers of English, pp. 101-132.

Carroll, J.B. (1990) Cognitive abilities in foreign language aptitude: Then and now. In *Language Aptitude Reconsidered*, eds. T. Parry and C. Stansfield, New Jersey: Prentice Hall Regents, pp. 11-29.

Carroll, J.B. and Sapon, S.M. (1959) *Modern Languages Aptitude Test.* New York: Psychological Corporation.

Casanave, C. (2009) Training for writing or training for reality? Challenges facing EFL writing teachers and students in language teacher education programs. In *Writing in Foreign Language Contexts: Learning, Teaching, and Research*, ed. R. Manchón, Electronic ed. Bristol: Multilingual Matters, pp. 256-277.

Chakrabarti, R. (2013) *South Korea's schools: Long days, high results.* Online ed. http://www.bbc.co.uk/news/education-25187993: BBC News.

Chandler, J. (2003) The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing,* 12 (3), pp. 267-296.

Chanock, K. (2000) Comments on essays: Do students understand what tutors write? *Teaching in Higher Education,* 5 (1), pp. 95-105.

Chaudron, C. (1977) A descriptive model of discourse in the corrective treatment of learners' errors. *Language Learning,* 27 (1), pp. 29-46.

Chen, P. (2016) Learners' metalinguistic and affective performance in blogging to write. *Computer Assisted Language Learning,* 29 (4), pp. 790-814.

Cheng, X. (2000) Asian students' reticence revisited. *System,* 28 (3), pp. 435-446.

Chesney, T. and Su, D.K.S. (2010) The impact of anonymity on weblog credibility. *International Journal of Human-Computer Studies,* 68 (10), pp. 710-718.

Chi, M.T.H., De Leeuw, N., Chiu, M. and Lavancher, C. (1994) Eliciting self-explanations improves understanding. *Cognitive Science,* 18 (3), pp. 439-477.

Chun, A.E., Day, R.R., Chenoweth, N.A. and Luppescu, S. (1982) Errors, interaction, and correction: A study of native-nonnative conversations. *TESOL Quarterly,* 16 (4), pp. 537-547.

CIA (2015) *The World Factbook: Languages.* Central Intelligence Agency. [Online]. Available from: <https://www.cia.gov/library/publications/the-world-factbook/fields/2098.html#ks> (Accessed 15 April, 2015).

Cochran, W.G. (1947) Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics,* 3 (1), pp. 22-38.

Cohen, L. and Holliday, M. (1982) *Statistics for Social Scientists.* London: Harper & Row.

Cohen, L., Towbes, L. and Flocco, R. (1988) Effects of induced mood on self-reported life events and perceived and received social support. *Journal of Personality & Social Psychology,* 55 (4), pp. 669-674.

Cohen, L., Manion, L. and Morrison, K. (2011) *Research Methods in Education.* Seventh ed. New York: Routledge.

Cohen, J.D., Servan-Schreiber, D. and McClelland, J.L. (1992) A parallel distributed processing approach to automaticity. *The American Journal of Psychology,* 105 (2), pp. 239-269.

CollegeBoard (2015) *Accuplacer.* The College Board. [Online]. Available from: <https://accuplacer.collegeboard.org/?CSRF=D5LG-0DTZ-8T07-JAW2-GA5V-VC5S-TV14-XE1S&affiliateId=rdr&bannerId=accuplacer> (Accessed 22 January, 2015).

Cortina, J.M. (1993) What Is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology,* 78 (1), pp. 98-104.

Craik, F.I.M. and Lockhart, R.S. (1972) Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior,* 11 (6), pp. 671-684.

Cramer, R., Neal, T. and Brodsky, S. (2009) Self-efficacy and confidence: Theoretical distinctions and implications for trial consultation. *Consulting Psychology Journal: Practice & Research,* 61 (4), pp. 319-334.

Crosthwaite, P. (in press) Does EAP writing instruction reduce L2 errors? Evidence from a longitudinal corpus of L2 EAP essays and reports. *International Review of Applied Linguistics in Language Teaching,* pp. 1-29.

Cumming, A. (1990) Metalinguistic and ideational thinking in second language composing. *Written Communication,* 7 (4), pp. 482-511.

Cumming, A. and Riazi, A. (2000) Building models of adult second-language writing instruction. *Learning and Instruction,* 10 (1), pp. 55-71.

Cummins, J. (1978) Bilingualism and the development of metalinguistic awareness. *Journal of Cross-Cultural Psychology,* 9 (2), pp. 131-149.

Datta, L. (2006) Paradigm wars: A basis for peaceful coexistence and beyond. In *Mixed Methods*, ed. A. Bryman, London: SAGE, pp. 31-52.

de Bot, K. (1996) The psycholinguistics of the output hypothesis. *Language Learning,* 46 (3), pp. 529-555.

de Jong, N. and Perfetti, C. (2011) Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning,* 61 (2), pp. 533-568.

DeKeyser, R.M. (1993) The effect of error correction on L2 grammar knowledge and oral proficiency. *The Modern Language Journal,* 77 (4), pp. 501.

DeKeyser, R. (1994) Implicit and explicit learning of L2 grammar: A pilot study. *TESOL Quarterly,* 28 (1), pp. 188-194.

DeKeyser, R. (1995) Learning second language grammar rules. *Studies in Second Language Acquisition,* 17 (3), pp. 379-410.

DeKeyser, R. (1998) Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In *Focus on Form in Classroom Second Language Acquisition.*, eds. J. Williams and C. Doughty, Cambridge: Cambridge University Press, pp. 42-63.

DeKeyser, R. (2007a) Introduction: Situating the concept of practice. In *Practice in a Second Language: Perspectives from applied linguistics and cognitive psychology.*, ed. R. DeKeyser, Cambridge: Cambridge University Press, pp. 1-18.

DeKeyser, R.M. (2007b) Skill acquisition theory. In *Theories in Second Language Acquisition*, eds. B. VanPatten and J. Williams, London: Lawrence Erlbaum Associates, pp. 97-113.

Delanty, G. and Strydom, P. (2003) Pragmatism, semiotics and transcendental pragmatism: Introduction. In *Philosophies of Social Science*, eds. G. Delanty and P. Strydom, Philadelphia: Open University Press, pp. 277-285.

Derrick, D. (2016) Instrument reporting practices in second language research. *TESOL Quarterly,* 50 (1), pp. 132–153.

Dewey, J. (2011) The need for a recovery of philosophy. In *The Pragmatism Reader*, eds. R. Talisse and S. Aikin, Princeton: Princeton University Press, pp. 109-154.

Dindyal, J., Tay, E.G., Quek, K.S., Leong, Y.H., Toh, T.L., Toh, P.C. and Ho, F.H. (2014) Cognitive scaffolding for problem solving: Use of the practical worksheet. *Proceedings of the Problem@Web International Conference: Technology, creativity and affect in mathematical problem solving*, eds. S. Carreira, N. Amado, K. Jones & H. Jacinto, Faro, Portugal: Universidade do Algarve, pp. 265.

Donato, R. (1994) Collective scaffolding in second language learning. In *Vygotskian Approaches to Second Language Research*, eds. J. Lantolf and G. Appel, London: Ablex Publishing, pp. 33-56.

Donnell, L.A. and Gettinger, M. (2015) Elementary school teachers' acceptability of school reform: Contribution of belief congruence, self-efficacy, and professional development. *Teaching and Teacher Education,* 51, pp. 47-57.

Dörnyei, Z. and Kormos, J. (2000) The role of individual and social variables in oral task performance. *Language Teaching Research,* 4 (3), pp. 275-300.

Dörnyei, Z. (2007) *Research Methods in Applied Linguistics.* Oxford University Press, USA.

Dörnyei, Z. and Ryan, S. (2015) *Psychology of the Language Learner Revisited.* Electronic ed. New York: Routledge.

Doughty, C. and Varela, E. (1998) Communicative focus on form. In *Focus on Form in Classroom Second Language Acquisition.*, eds. C. Doughty and J. Williams, Cambridge: Cambridge University Press, pp. 114-138.

Doughty, C. and Williams, J. (1998) Pedagogical choices in focus on form. In *Focus on Form in Classroom Second Language Acquisition.*, eds. C. Doughty and J. Williams, Cambridge: Cambridge University Press, pp. 197-261.

Doughty, C. (2005) Instructed SLA: Constraints, compensation, and enhancement. In *The Handbook of Second Language Acquisition.*, eds. C. Doughty and M.H. Long, Electronic ed. Oxford: Blackwell (no pages)

Dulay, H., Burt, M. and Krashen, S. (1982) *Language Two.* Oxford: Oxford University Press.

Educational Testing Service (2016) *TOEFL® Tips.* ETS TOEFL. [Online]. Available from: <https://www.ets.org/toefl/ibt/prepare/tips/> (Accessed December 5, 2016).

EF EPI (2017) *EF English Proficiency Index (EPI).* EF Education First. [Online]. Available from: <http://www.ef.co.uk/epi/regions/asia/south-korea/> (Accessed 12 November, 2017).

Ekholm, E., Zumbrunn, S. and Conklin, S. (2015) The relation of college student self-efficacy toward writing and writing self-regulation aptitude: Writing feedback perceptions as a mediating variable. *Teaching in Higher Education,* 20 (2), pp. 197-207.

Elder, C. (2009) Validating a test of metalinguistic knowledge. In *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*, eds. R. Ellis S. Loewen C. Elder R. Erlam J. Philp and H. Reinders, Electronic ed. Bristol: Multilingual Matters, pp. 113-138.

Ellis, R. (1993) The structural syllabus and second language acquisition. *TESOL Quarterly,* 27 (1), pp. 91-113.

Ellis, R. (2001) Introduction: Investigating form-focused instruction. *Language Learning,* 51 (s1), pp. 1-46.

Ellis, R. (2004) The definition and measurement of L2 explicit knowledge. *Language Learning,* 54 (2), pp. 227-275.

Ellis, R. (2005) Measuring implicit and explicit knowledge of a second language - A psychometric study. *Studies in Second Language Acquisition,* 27 (2), pp. 141-172.

Ellis, R. (2006) Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly,* 40 (1), pp. 83-107.

Ellis, R. (2009a) Implicit and explicit learning, knowledge, and instruction. In *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*, eds. R. Ellis S. Loewen C. Elder R. Erlam J. Philp and H. Reinders, Electronic ed. Bristol: Multilingual Matters, pp. 3-29.

Ellis, R. (2009b) Corrective feedback and teacher development. *L2 Journal,* [Online]. 1 , pp. 14 June 2017-3-18. Available from: <http://escholarship.org/uc/item/2504d6w3#> (Accessed 14 June 2017).

Ellis, R. (2009c) A typology of written corrective feedback types. *ELT Journal,* 63 (2), pp. 97-107.

Ellis, R. (2010) A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition,* 32, pp. 335-349.

Ellis, R., Sheen, Y., Murakami, M. and Takashima, H. (2008) The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System,* 36 (3), pp. 353-371.

Elwood, J.A. and Bode, J. (2014) Student preferences vis-à-vis teacher feedback in university EFL writing classes in Japan. *System,* 42, pp. 333-343.

Erlam, R. (2005) Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research,* 9 (2), pp. 147-171.

Esteban, N. and Roca de Larios, J. (2010) The use of models as a form of written corrective feedback to secondary school pupils of English. *International Journal of English Studies,* 10 (2), pp. 143-170.

Evans, N.W., Hartshorn, K.J., McCollum, R.M. and Wolfersberger, M. (2010) Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research,* 14 (4), pp. 445-463.

Evans, S. and Morrison, B. (2011) The first term at university: Implications for EAP. *ELT Journal,* 65 (4), pp. 387-397.

Evans, Hartshorn and Strong-Krause (2011) The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System,* 39 (2), pp. 229-239.

Ferguson, C. (2009) An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice,* 40 (5), pp. 532-538.

Ferris, D.R. (1995) Student reactions to teacher response in multiple-draft composition classrooms. *TESOL Quarterly,* 29 (1), pp. 33-53.

Ferris, D. (1999) The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing,* 8 (1), pp. 1-11.

Ferris, D.R. (2004) The "grammar correction" debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime …?). *Journal of Second Language Writing,* 13 (1), pp. 49-62.

Ferris, D. (2006) Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In *Feedback in Second Language Writing: Contexts and issues*, eds. K. Hyland and F. Hyland, Cambridge: Cambridge University Press, pp. 81-104.

Ferris, D. (2009) *Teaching College Writing to Diverse Student Populations.* Ann Arbor: The University of Michigan Press.

Ferris, D.R. (2010) Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition,* 32 (02), pp. 181-201.

Ferris, D., Brown, J., Liu, H. and Stine, M. (2011) Responding to L2 students in college writing classes: Teacher perspectives. *TESOL Quarterly,* 45 (2), pp. 207-234.

Finn, B. and Metcalfe, J. (2010) Scaffolding feedback to maximize long-term error correction. *Memory & Cognition,* 38 (7), pp. 951-61.

Fisher, R., Cavanagh, J. and Bowles, A. (2011) Assisting transition to university: using assessment as a formative learning tool. *Assessment & Evaluation in Higher Education,* 36 (2), pp. 225-237.

Fitts, P. (1964) Perceptual-motor skill learning. In *Categories of Human Learning.*, ed. A.W. Melton, New York: Academic Press, pp. 244-285.

Flanders, N. (1970) *Analyzing Teaching Behavior.* London: Addison-Wesley Publishing Company.

Flick, U. (2014) *An Introduction to Qualitative Research.* Fifth ed. London: Sage.

Fulcher, G. (2010) *Practical Language Testing.* London: Hodder Education.

Gage, N.L. (1989) The paradigm wars and their aftermath: A "historical" sketch of research on teaching since1989. *Educational Researcher,* 18 (7), pp. 4-10.

Gaies, S.J. (1980) T-Unit analysis in second language research: Applications, problems and limitations. *TESOL Quarterly,* 14 (1), pp. 53-60.

Galloway, N. and Rose, H. (2015) *Introducing Global Englishes.* Electronic ed. Abingdon: Routledge.

Gass, S. (2005) Input and interaction. In *The Handbook of Second Language Acquisition.*, eds. C. Doughty and M.H. Long, Electronic ed. Oxford: Blackwell Reference Online (no pages).

Gass, S.and Mackey, A. (2007) Input, interaction, and output in second language acquisition. In *Theories in Second Language Acquisition*, eds. B. VanPatten and J. Williams, London: Lawrence Erlbaum Associates, pp. 175-199.

Gass, S. (2018) *Input, Interaction, and the Second Language Learner.* Kindle ed. New York: Routledge.

Gass, S., Svetics, I. and Lemelin, S. (2003) Differential effects of attention. *Language Learning,* 53 (3), pp. 497-546.

Goldstein, L. (2016) Making use of teacher written feedback. In *Handbook of Second and Foreign Language Writing*, eds. R. Manchón and P.K. Matsuda, Electronic ed. Berlin: De Gruyter, pp. 407-430.

González-Becerra, I. (in press) Language learning among STEM students: motivational profile and attitudes of undergraduates in a UK institution. *The Language Learning Journal,* pp. 1-17.

Grabowski, J. (1996) Writing and speaking: Common grounds and differences toward a regulation theory of written language production. In *The Science of Writing: Theories, Methods, Individual Differences and Applications*, eds. C.M. Levy and S. Ransdell, Mahwah: Lawarence Erlbaum, pp. 73-91.

Graham, S.E. and Harris, K.R. (2014) Conducting high quality writing intervention research: Twelve recommendations. *Journal of Writing Research,* 6 (2), pp. 89-123.

Graham, S. and Hebert, M. (2011) Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review,* 81 (4), pp. 710-744.

Granott, N. (2005) Scaffolding dynamically toward change: Previous and new perspectives. *New Ideas in Psychology*, 23 (3), pp. 140-151.

Gravetter, F. and Forzano, L. (2016) *Research Methods for the Behavioral Sciences.* 5th ed. Stamford: Cengage Learning.

Grbich, C. (2013) *Qualitative Data Analysis: An Introduction.* Second ed. London: SAGE Publications Ltd.

Greasley, P. and Cassidy, A. (2009) When it comes round to marking assignments: how to impress and how to 'distress' lecturers ... *Assessment & Evaluation in Higher Education,* 35 (2), pp. 173-189.

Gresham, F., MacMillan, D., Beebe-Frankenberger, M. and Bocian, K. (2000) Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice,* 15 (4), pp. 198-205.

Griggs, P., Carol, R. and Bange, P. (2002) La dimension cognitive dans l'apprentissage des langues étrangères. *Revue française de linguistique appliquée,* VII (2), pp. 25-38.

Grix, J. (2002) Introducing students to the generic terminology of social research. *Politics,* 22 (3), pp. 175-186.

Gu, Y. (2002) Gender, academic major, and vocabulary learning strategies of Chinese EFL learners. *RELC Journal,* 33 (1), pp. 35-54.

Guilloteaux, M. (2013) Motivational strategies for the language classroom: Perceptions of Korean secondary school English teachers. *System,* 41 (1), pp. 3-14.

Gutiérrez, X. (2013) Metalinguistic knowledge, metalingual knowledge, and proficiency in L2 Spanish. *Language Awareness,* 22 (2), pp. 176-191.

Gutiérrez, X. (2016) Analyzed knowledge, metalanguage, and second language proficiency. *System,* 60, pp. 42-54.

Hamel, H. (1994) *Hamel's Journal and a Description of the Kingdom of Korea 1653-1666 (Translated from the Dutch manuscript by Br. Jean-Paul Buys, of Taizé).* Seoul: The Royal Asiatic Society Korea Branch; Seoul Press.

Hammersley, M. (2006) The paradigm wars: Reports from the front. In *Mixed Methods*, ed. A. Bryman, London: SAGE, pp. 3-19.

Hammersley, M. (2003) Recent radical criticism of interview studies: Any implications for the sociology of education? *British Journal of Sociology of Education,* 24 (1), pp. 119-126.

Han, M.-. (2005) *딸애 교실에서 '야자' 하고 왔어요.* Online ed. http://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=103&oid=047&aid=0000064376: OhmyNews.

Han, Y. and Ellis, R. (1998) Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research,* 2 (1), pp. 1-23.

Han, Z. and Finneran, R. (2014) Re-engaging the interface debate: strong, weak, none, or all? *International Journal of Applied Linguistics,* 24 (3), pp. 370-389.

Hanaoka, O. (2007) Output, noticing, and learning: An investigation into the role of spontaneous attention to form in a four-stage writing task. *Language Teaching Research,* 11 (4), pp. 459-479.

Hanaoka, O. and Izumi, S. (2012) Noticing and uptake: Addressing pre-articulated covert problems in L2 writing. *Journal of Second Language Writing,* 21 (4), pp. 332-347.

Hardman, J. (2016) Tutor-student interaction in seminar teaching: Implications for professional development. *Active Learning in Higher Education,* 17 (1), pp. 63-76.

Hartshorn, K., Evans, N., Merrill, P., Sudweeks, R., Strong-Krause, D. and Anderson, N. (2010) Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly,* 44 (1), pp. 84-109.

Hau, K. and Salili, F. (1991) Structure and semantic differential placement of specific causes: Academic causal attributions by Chinese students in Hong Kong. *International Journal of Psychology,* 26 (2), pp. 175-193.

Hedgcock, J. and Lefkowitz, N. (1996) Some input on input: Two analyses of student response to expert feedback in L2 writing. *The Modern Language Journal,* 80 (3), pp. 287-308.

Hedge, T. (2000) *Teaching and Learning in the Language Classroom.* Oxford: Oxford University Press.

Higgins, R., Hartley, P. and Skelton, A. (2001) Getting the message across: The problem of communicating assessment feedback. *Teaching in Higher Education,* 6 (2), pp. 269-274.

Ho, E.S. (2009) Characteristics of East Asian learners: What we learned from PISA. *Educational Research Journal,* 24 (2), pp. 327-348.

Howe, K.R. (1988) Against the Quantitative-Qualitative Incompatibility Thesis or Dogmas Die Hard. *Educational Researcher,* 17 (8), pp. 10-16.

Hsieh, P. and Kang, H. (2010) Attribution and self-efficacy and their interrelationship in the Korean EFL context. *Language Learning,* 60 (3), pp. 606-627.

Hu, G. (2011) Metalinguistic knowledge, metalanguage, and their relationship in L2 learners. *System,* 39 (1), pp. 63-77.

Hunt, K. (1965) *Grammatical structures written at three grade levels: NCTE research report no. 3.* Champaign, IL: National Council of Teachers of English. <http://files.eric.ed.gov/fulltext/ED113735.pdf>.

Hyland, F. (1998) The impact of teacher written feedback on individual writers. *Journal of Second Language Writing,* 7 (3), pp. 255-286.

Hyland, K. (2013a) Faculty feedback: Perceptions and practices in L2 disciplinary writing. *Journal of Second Language Writing,* 22, pp. 240-253.

Hyland, K. (2013b) Student perceptions of hidden messages in teacher written feedback. *Studies in Educational Evaluation,* 39 (3), pp. 180-187.

Impara, J. and Plake, B. (1998) Teachers' ability to estimate item difficulty: A test of the assumptions in the Angoff standard setting method. *Journal of Educational Measurement,* 35 (1), pp. 69-81.

Ishikawa, M. (2013) Examining the effect of written languaging: the role of metanotes as a mediator of second language learning. *Language Awareness,* 22 (3), pp. 220-233.

Ishikawa, M. and Suzuki, W. (2016) The effect of written languaging on learning the hypothetical conditional in English. *System*, 58, pp. 97-111.

Iwashita, N. (2003) Negative feedback and positive evidence in task-based interaction: Differential effects on L2 development. *Studies in Second Language Acquisition,* 25 (1), pp. 1-36.

Jager, W. (2003) Breaking 'bad habits': A dynamical perspective on habit formation and change. In *Human Decision Making and Environmental Perception: Understanding and Assisting Human*

*Decision Making in Real-life Settings.*, eds. L. Hendrickx W. Jager and L. Steg, Groningen: University of Groningen.

James, W. (1890) *Principles of Psychology, V.1.* London: Macmillan and Co.

Jang, Y. (2006) 침묵의 유형과 교육적 의의 (The types of silence and it's [sic] educational meaning). *한국화법학회,* 9, pp. 43-68.

Johnson, E., Mellard, D., Fuchs, D. and McKnight, M. (2006) *Responsiveness to intervention (RTI): How to do it.* Lawrence, KS: National Research Center on Learning Disabilities.

Johnson, K. (1982) *Communicative Syllabus Design and Methodology.* Oxford: Pergamon Press.

Johnson, E.J., Tuskenis, A.D., Howell, G.L. and Jaroszewski, K. (2011) Development and effects of a writing and thinking course in psychology. *Teaching of Psychology,* 38 (4), pp. 229-236.

Johnson, R.B., Onwuegbuzie, A.J. and Turner, L.A. (2007) Toward a definition of mixed methods research. *Journal of Mixed Methods Research,* 1 (2), pp. 112-133.

Johnston, A. and Lange, C. (2006) The beginnings of standardization - An epilogue. In *The Beginnings of Standardization: Language and Culture in Fourteenth-Century England*, ed. U. Schaefer, Frankfurt: Peter Lang, pp. 183-196.

Joo, Y.J., Lim, K.Y. and Kim, J. (2013) Locus of control, self-efficacy, and task value as predictors of learning outcome in an online university context. *Computers & Education,* 62, pp. 149-158.

Jordan, G. (2004) *Theory Construction in Second Language Acquisition.* Electronic ed. Philadelphia: John Benjamins Publishing Company.

Junqueira, L. and Kim, Y. (2013) Exploring the relationship between training, beliefs, and teachers' corrective feedback practices: A case study of a novice and an experienced ESL teacher. *The Canadian Modern Language Journal,* 69 (2), pp. 181-206.

Kachru, B.B. (1988) The sacred cows of English. *English Today,* 16, pp. 3-8.

Kalaja, P.and Barcelos, A.M.F. (2006) Introduction: Rationale for 'Beliefs about SLA: New research approaches'. In *Beliefs About SLA: New research approaches*, eds. P. Kalaja and A.M.F. Barcelos, New York: Springer, pp. 1-4.

Kang, S. (2010) *Korea is heaven for illegal private English tutoring.* Online ed. http://www.koreatimes.co.kr/www/news/nation/2010/02/117_60083.html: The Korean Times.

Kang, Y. and Pyun, D. (2013) Mediation strategies in L2 writing processes: A case study of two Korean language learners. *Language, Culture and Curriculum,* 26 (1), pp. 52-67.

Kayi-Aydar, H. (2013) Scaffolding language learning in an academic ESL classroom. *ELT Journal,* 67 (3), pp. 324-335.

Kim, U. and Park, Y.S. (2006) Indigenous psychological analysis of academic achievement in Korea: The influence of self-efficacy, parents, and culture. *International Journal of Psychology,* 41 (4), pp. 287-292.

Kim, J.W., Ritter, F.E. and Koubek, R.J. (2013) An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science,* 14 (1), pp. 22-37.

Kim, S.H.O. and Elder, C. (2005) Language choices and pedagogic functions in the foreign language classroom: A cross-linguistic functional analysis of teacher talk. *Language Teaching Research,* 9 (4), pp. 355-380.

Kim, S.H.O. and Elder, C. (2008) Target language use in foreign language classrooms: Practices and perceptions of two native speaker teachers in New Zealand. *Language, Culture and Curriculum,* 21 (2), pp. 167-185.

Klapper, J. and Rees, J. (2003) Reviewing the case for explicit grammar instruction in the university foreign language learning context. *Language Teaching Research,* 7 (3), pp. 285-314.

Klenowski, V. (1995) Student self-evaluation processes in student-centred teaching and learning contexts of Australia and England. *Assessment in Education: Principles, Policy & Practice,* 2 (2), pp. 145-163.

Knouzi, I., Swain, M., Lapkin, S. and Brooks, L. (2010) Self-scaffolding mediated by languaging: microgenetic analysis of high and low performers. *International Journal of Applied Linguistics,* 20 (1), pp. 23-49.

Ko, J., Schallert, D. and Walters, K. (2003) Rethinking scaffolding: Examining negotiation of meaning in an ESL storytelling task. *TESOL Quarterly,* 37 (2), pp. 303-324.

Korea University (2015) *고려대학교 입학처.* Admissions Office of Korea University. [Online]. Available from: <http://oku.korea.ac.kr/admissions/index.oku> (Accessed 18 April, 2015).

Kormos, J. (2012) The role of individual differences in L2 writing. *Journal of Second Language Writing,* 21 (4), pp. 390-403.

Koro-Ljungberg, M. (2004) Impossibilities of reconciliation: Validity in mixed theory projects. *Qualitative Inquiry,* 10 (4), pp. 601-621.

Koro-Ljungberg, M., Yendol-Hoppey, D., Smith, J.J. and Hayes, S.B. (2009) (E)pistemological awareness, instantiation of methods, and uninformed methodological ambiguity in qualitative research projects. *Educational Researcher,* 38 (9), pp. 687-699.

Koro-Ljungberg, M., Douglas, E.P., Therriault, D., Malcolm, Z. and McNeill, N. (2013) Reconceptualizing and decentering think-aloud methodology in qualitative research. *Qualitative Research,* 13 (6), pp. 735-753.

KOSTAT (2015a) *2014 년 4/4 분기 및 연간 가계동향.* http://kostat.go.kr/portal/korea/index.action: 통계청 (Statistics Korea).

KOSTAT (2015b) *2014 년 사교육비조사 결과.* http://kostat.go.kr/portal/korea/index.action: 통계청 (Statistics Korea).

Krashen, S. (1982) *Principles and Practice in Second Language Acquisition.* Oxford: Pergamon.

Kruger, J. and Dunning, D. (1999) Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality & Social Psychology,* 77 (6), pp. 1121-1134.

KTO (2012) *한국방문의 해 위원회 - 숨기편.* Korea Tourism Organisation. [Online]. Available from: <http://youtu.be/tT60Gw7wkuA> (Accessed 1 August, 2013).

Kuckartz, U. (2014) *Qualitative Text Analysis: A Guide to Methods, Practice and Using Software.* Electronic ed. London: Sage.

Kuder, G. and Richardson, M. (1937) The theory of the estimation of test reliability. *Psychometrika,* 2 (3), pp. 151-160.

Kuiken, F. and Vedder, I. (2002) The effect of interaction in acquiring the grammar of a second language. *International Journal of Educational Research,* 37 (3–4), pp. 343-358.

Kwon, S.K., Lee, M. and Shin, D. (2017) Educational assessment in the Republic of Korea: Lights and shadows of high-stake exam-based education system. *Assessment in Education: Principles, Policy & Practice,* 24 (1), pp. 60-77.

Langslow, D. (1995) *The Colleges of Oxford University 'Mode N' Test for Entrance in Classics 1995: Language Aptitude Test.* Oxford: Oxford University.

Langslow, D. (2014) *Personal communication.*

Lapkin, S., Swain, M. and Knouzi, I. (2008) French as a second language students learn the grammatical concept of voice: Study design, materials development, and pilot data. In *Sociocultural Theory and the Teaching of Second Languages*, eds. J. Lantolf and M. Poehner, London: Equinox Press, pp. 228-255.

Lee, G. (2009) Speaking up: Six Korean students' oral participation in class discussions in US graduate seminars. *English for Specific Purposes,* 28 (3), pp. 142-156.

Lee, G. and Schallert, D.L. (2008) Meeting in the margins: Effects of the teacher–student relationship on revision processes of EFL college students taking a composition course. *Journal of Second Language Writing,* 17 (3), pp. 165-182.

Lee, I. (2004) Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing,* 13 (4), pp. 285-312.

Lee, I. (2008) Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing,* 17 (3), pp. 144-164.

Lee, I., Mak, P. and Burns, A. (2016) EFL teachers' attempts at feedback innovation in the writing classroom. *Language Teaching Research,* 20 (2), pp. 248-269.

Lee, J., Bong, M. and Kim, S. (2014) Interaction between task values and self-efficacy on maladaptive achievement strategy use. *Educational Psychology,* 34 (5), pp. 538-560.

Lee, L. (2008) Focus-on-form through collaborative scaffolding in expert-to-novice online interaction. *Language Learning & Technology,* 12 (3), pp. 53-72.

Lee, W.Y. (2011) *Test takers undergo stressful rite of passage.* English edition ed. Seoul: The Korea Herald. Available from http://www.koreaherald.com/view.php?ud=20111110000751. (Last accessed 10 February 2013)

Leech, N. and Onwuegbuzie, A. (2007) An array of qualitative data analysis tools: A call for data analysis triangulation. *School Psychology Quarterly,* 22 (4), pp. 557-584.

Leeman, J. (2007) Feedback in L2 learning: Responding to errors during practice. In *Practice in a Second Language: Perspectives from applied linguistics and cognitive psychology*, ed. R. DeKeyser, Cambridge: Cambridge University Press, pp. 111-137.

Leow, R.P. (2015) *Explicit Learning in the L2 Classroom: A student-centered approach.* New York: Routledge.

Lepper, M., Drake, M. and O'Donnell-Johnson, T. (1997) Scaffolding techniques of expert human tutors. In *Scaffolding Student Learning: Instructional approaches and issues*, eds. K. Hogan and M. Pressley, Cambridge, MA: Brookline Books/Lumen Editions, pp. 108-144.

Levine, T. and Hullett, C. (2002) Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research,* 28 (4), pp. 612-625.

Li, A. and Post, B. (2015) *Personal communication.*

Li, A. and Post, B. (2014) L2 acquisition of prosodic properties of speech rhythm. *Studies in Second Language Acquisition,* 36 (2), pp. 223-255.

Li, S. (2009) The differential effects of implicit and explicit feedback on second language (L2) learners at different proficiency levels. *Applied Language Learning,* 19 (1,2), pp. 53-78.

Li, S. (2015) The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics,* 36 (3), pp. 385-408.

Li, S. (2013) The interactions between the effects of implicit and explicit feedback and individual differences in language analytic ability and working memory. *The Modern Language Journal,* 97 (3), pp. 634-654.

Li, X. (2013) *Numeral Classifiers in Chinese: The Syntax-Semantics Interface.* Electronic ed. Berlin: De Gruyter.

Lightbown, P. and Spada, N. (2013) *How Languages are Learned.* 4th ed. Oxford: Oxford University Press.

Lincoln, Y. and Guba, E. (1985) *Naturalistic Inquiry.* London: Sage.

Lindseth, M. (2016) The effects of form-focused instruction on the acquisition of subject-verb inversion in German. *Foreign Language Annals,* 49 (1), pp. 10-22.

Littleton, K. (2013) Adaptation and authority in scaffolding and teacher–student relationships: Commentary on the Special Issue 'Conceptualising and grounding scaffolding in complex educational contexts'*. Learning, Culture and Social Interaction*, 2 (1), pp. 52-55.

Liu, D. and Gleason, J.L. (2002) Acquisition of the article *the* by non-native speakers of English: An analysis of four nongeneric uses. *Studies in Second Language Acquisition,* 24 (1), pp. 1-26.

Lo, Y.Y. and Macaro, E. (2015) Getting used to content and language integrated learning: what can classroom interaction reveal? *The Language Learning Journal,* 43 (3), pp. 239-255.

Long, M.H. (1981) Input, interaction and second language acquisition. In *Native Language and Foreign Language Acquisition. Annals of the New York Academy of Sciences Vol. 379*, ed. H. Winitz, pp. 259-278.

Long, M. (1991) Focus on form: A design feature in language teaching methodology. In *Foreign Language Research in Cross-Cultural Perspective.*, eds. R.B. Ginsberg and C. Kramsch, Electronic ed. Amsterdam: John Benjamins Publishing Company, pp. 39-52.

Long, M.H. (1996) The role of the linguistic environment in second language acquisition. In *Handbook of Second Language Acquisition.*, eds. W.C. Ritchie and T.K. Bhatia, San Diego: Academic Press, pp. 413-468.

Long, M. (2017) Instructed second language acquisition (ISLA): Geopolitics, methodological issues, and some major research questions. *Instructed Second Language Acquisition,* 1 (1), pp. 7-44.

Lyster, R. and Ranta, L. (1997) Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition,* 19 (1), pp. 37-66.

Lyster, R. (2004) Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition,* 26 (3), pp. 399-432.

Lyster, R. and Sato, M. (2013) Skill acquisition theory and the role of practice in L2 development. In *Contemporary Approaches to Second Language Acquisition.*, eds. M.M.D.P. García M.M.J. Gutierrez and A.M. Martínez, Electronic ed. Amsterdam: John Benjamins Publishing Company, pp. 71-91.

Mackey, A. and Gass, S. (2016) *Second Language Research: Methodology and Design.* E-book; Second ed. New York: Routledge.

MacKey, A. and Philp, J. (1998) Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal,* 82 (3), pp. 338-356.

Manchón, R. (2011) Situating the learning-to-write and writing-to-learn dimensions of L2 writing. In *Learning-to-Write and Writing-to-Learn in an Additional Language*, ed. R. Manchón, Amsterdam: John Benjamins, pp. 3-14.

Marlina, R. (2009) “I don’t talk or I decide not to talk? Is it my culture?”—International students’ experiences of tutorial participation. *International Journal of Educational Research,* 48 (4), pp.235-244

Mascolo, M.F. (2005) Change processes in development: The concept of coactive scaffolding*. New Ideas in Psychology*, 23 (3), pp. 185-196.

Master, P. (1995) Consciousness raising and article pedagogy. In *Academic Writing in a Second Language*, eds. D. Belcher and G. Brain, Norwood: Ablex, pp. 103-204.

Mattern, K. and Packman, S. (2009) *Predictive validity of ACCUPLACER® scores for course placement: A meta-analysis.* http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2009-2-predictive-validity-accuplacer-scores-course-placement.pdf: Accuplacer.

May, H., Perez-Johnson, I., Haimson, J., Sattar, S. and Gleason, P. (2009) *Using state tests in education experiments: A discussion of the issues.* U.S. Department of Education: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences.

McCarthy, M. (1991) *Discourse Analysis for Language Teachers.* Cambridge: Cambridge University Press.

McLaughlin, B., Rossman, T. and McLeod, B. (1983) Second language learning: An information-processing perspective. *Language Learning,* 33 (2), pp. 135-158.

McLean, A.J., Bond, C.H. and Nicholson, H.D. (2015) An anatomy of feedback: A phenomenographic investigation of undergraduate students' conceptions of feedback. *Studies in Higher Education*, 40 (5), pp. 921-932.

McMartin-Miller, C. (2014) How much feedback is enough?: Instructor practices and student attitudes toward error treatment in second language writing. *Assessing Writing,* 19, pp. 24-35.

Meara, P. (2005) *LLAMA Swansea Language Aptitude Tests.* Swansea: University of Wales Swansea. [Online]. Available from: <http://www.lognostics.co.uk/tools/llama/index.htm> (Accessed 20 January, 2014).

Miao, P. and Heining-Boynton, A. (2011) Initiation/ response/ follow-Up, and response to intervention: Combining two models to improve teacher and student performance. *Foreign Language Annals,* 44 (1), pp. 65-79.

Miller, G.A. (1956) The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review,* 63 (2), pp. 81-97.

Mirzaei, A. and Eslami, Z.R. (2015) ZPD-activated languaging and collaborative L2 writing. *Educational Psychology,* 35 (1), pp. 5-25.

Moodie, I. (2016) The anti-apprenticeship of observation: How negative prior language learning experience influences English language teachers’ beliefs and practices. *System,* 60, pp. 29-41.

Moradian, M.R., Miri, M. and Hossein Nasab, M. (2017) Contribution of written languaging to enhancing the efficiency of written corrective feedback. *International Journal of Applied Linguistics*, 27 (2), pp. 406–426.

Morony, S., Kleitman, S., Lee, Y.P. and Stankov, L. (2013) Predicting achievement: Confidence vs self-efficacy, anxiety, and self-concept in Confucian and European countries*. International Journal of Educational Research*, 58, pp. 79-96.

Mortimer, E. and Scott, P. (2003) *Meaning Making in Secondary Science Classrooms.* Electronic ed. Maidenhead: McGraw-Hill International (UK) Ltd.

Mulliner, E. and Tucker, M. (2017) Feedback on feedback practice: perceptions of students and academics. *Assessment & Evaluation in Higher Education,* 42 (2), pp. 266-288.

Muñoz, C. (2014) Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics,* 35 (4), pp. 463-482.

Muranoi, H. (2000) Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning,* 50 (4), pp. 617-673.

Murtagh, L. and Baker, N. (2009) Feedback to feed forward: Student response to tutors' written comments on assignments. *Practitioner Research in Higher Education,* 3 (1), pp. 20-28.

Myers, G. and Myers, M. (1992) *The Dynamics of Human Communication: A Laboratory Approach.* 6th ed. New York: McGraw-Hill, Inc.

Nassaji, H. and Swain, M. (2000) A Vygotskian perspective on corrective feedback in L2: The effect of random versus negotiated help on the learning of English articles. *Language Awareness,* 9 (1), pp. 34-51.

Nassaji, H. and Fotos, S. (2007) Issues in form-focused instruction and teacher education. In *Form-focused Instruction and Teacher Education: Studies in honour of Rod Ellis.*, eds. S. Fotos and H. Nassaji, Oxford: Oxford University Press, pp. 7-15.

Nassaji, H. (2011) Immediate learner repair and its relationship with learning targeted forms in dyadic interaction. *System,* 39 (1), pp. 17-29.

Nassaji, H. (2015) *Interactional Feedback Dimension in Instructed Second Language Learning.* Electronic ed. London: Bloomsbury Academic.

Nassaji, H. (2016) Anniversary article Interactional feedback in second language teaching and learning: A synthesis and analysis of current research. *Language Teaching Research,* 20 (4), pp. 535-562.

Nayar, P.B. (1997) ESL/EFL dichotomy today: Language politics or pragmatics? *TESOL Quarterly,* 31 (1), pp. 9-37.

Nazari, N. (2013) The effect of implicit and explicit grammar instruction on learners' achievements in receptive and productive modes. *Procedia - Social and Behavioral Sciences,* [Online]. 70, pp. 17 June 2017-156-162. Available from: <https://doi.org/10.1016/j.sbspro.2013.01.051> (Accessed 17 June 2017).

Negretti, R. (2012) Metacognition in student academic writing: A longitudinal study of metacognitive awareness and its relation to task perception, self-regulation, and evaluation of performance. *Written Communication,* 29 (2), pp. 142-179.

Ney, J. (1966) Review of *Differences in Grammatical Structures Written at Three Grade Levels*. *Language Learning,* 16 (3-4), pp. 230-235.

Nissani, M. and Hoefler-Nissani, D. (1992) Experimental studies of belief dependence of observations and of resistance to conceptual change. *Cognition and Instruction,* 9 (2), pp. 97-111.

Norris, J.M. and Manchón, R. (2012) Investigating L2 writing development from multiple perspectives. In *L2 Writing Development: Multiple Perspectives.*, ed. R. Manchón, Electronic ed. Berlin: De Gruyter Mouton, pp. 221-244.

Norris, J.M. and Ortega, L. (2000) Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning,* 50 (3), pp. 417-528.

Norris, J.M. and Ortega, L. (2001) Does type of instruction make a difference? Substantive findings from a meta-analytic review. In *Form-Focused Instruction and Second Language Learning*, ed. R. Ellis, Oxford: Blackwell Publishers, Ltd., pp. 157-213.

O'Donovan, B., Rust, C. and Price, M. (2016) A scholarly approach to solving the feedback dilemma in practice. *Assessment & Evaluation in Higher Education,* 41 (6), pp. 938-949.

O'Sullivan, M. (2004) The reconceptualisation of learner-centred approaches: A Namibian case study. *International Journal of Educational Development,* 24 (6), pp. 585-602.

Oakley, A. (2006) Paradigm wars: Some thoughts on a personal and public trajectory. In *Mixed Methods*, ed. A. Bryman, London: SAGE, pp. 21-30.

O'Donnell, C.L. (2008) Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research,* 78 (1), pp. 33-84.

OECD (2010) *PISA 2009 Results: What Students Know and Can Do - Student Performance in Reading, Mathematics and Science (Volume I).*

OECD (2013) Suicides. In *OECD Factbook 2013: Economic, Environmental and Social Statistics* http://www.oecd-ilibrary.org/sites/factbook-2013-en/12/01/03/index.html?itemId=/content/chapter/factbook-2013-97-en: OECD Publishing, pp. 236-237.

OECD (2014a) *Education at a Glance 2014: OECD Indicators.* OECD Publishing.

OECD (2014b) *Lessons from PISA for Korea.* <http://www.keepeek.com/Digital-Asset-Management/oecd/education/lessons-from-pisa-for-korea_9789264190672-en#page1>: OECD Publishing.

Oliver, R. (1995) Negative feedback in child NS-NNS conversation. *Studies in Second Language Acquisition,* 17 (4), pp. 459-481.

Oliver, R. and Mackey, A. (2003) Interactional context and feedback in child ESL classrooms. *The Modern Language Journal,* 87 (4), pp. 519-533.

Onwuegbuzie, A.J. and Leech, N.L. (2005) On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology,* 8 (5), pp. 375-387.

Oppenheim, A. (1992) *Questionnaire Design, Interviewing and Attitude Measurement.* Second ed. New York: Continuum.

Orsmond, P. and Merry, S. (2011) Feedback alignment: effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education,* 36 (2), pp. 125-136.

Ortega, L. (2009) Studying writing across EFL contexts: Looking back and moving forward. In *Writing in Foreign Language Contexts: Learning, Teaching, and Research.*, ed. R. Manchón, Electronic ed. Bristol: Multilingual Matters, pp. 232-255.

Ortega, L. (2012) Epilogue: Exploring L2 writing—SLA interfaces. *Journal of Second Language Writing,* 21 (4), pp. 404-415.

OUP (2014) *oxfordenglishtesting.com.* Oxford University Press. [Online]. Available from: <https://www.oxfordenglishtesting.com/defaultmrarticle.aspx?id=3074> (Accessed 9 September, 2015).

Pajares, M.F. (1992) Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research,* 62 (3), pp. 307-332.

Paradis, M. (2004) *A Neurolinguistic Theory of Bilingualism.* Electronic ed. Amsterdam: John Benjamins Publishing Company.

Park, E.S. (2013) Learner-generated noticing behavior by novice learners: Tracing the effects of learners' L1 on their emerging L2. *Applied Linguistics,* 34 (1), pp. 74-98.

Park, J. (2011) Metamorphosis of Confucian Heritage Culture and the possibility of an Asian education research methodology. *Comparative Education,* 47 (3), pp. 381-393.

Park, J. (2016) *[엠디팩트] 혀 길이 늘이는 설소대수술, 정말 영어 'R' 발음 좋아질까?* Online ed.

Parreño, A. (2015) Written corrective feedback impact on grammatical accuracy in L2 writing: A quantitative and qualitative look. *Writing and Pedagogy,* 7 (2-3), pp. 279-303.

Pawlak, M. and Tomczyk, E. (2013) Differential effects of input-providing and output-inducing corrective feedback on the acquisition of English passive voice. In *Language in Cognition and Affect.*, eds. E. Piechurska-Kuciel and E. Szymanska-Czaplak, New York: Springer, pp. 133-150.

Peacock, M. and Ho, B. (2003) Student language learning strategies across eight disciplines. *International journal of applied linguistics,* 13 (2), pp. 179-200.

Pica, T. (1994) Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning,* 44 (3), pp. 493-527.

Pimsleur, P. (2014) *Pimsleur Language Aptitude Battery.* Language Learning and Testing Foundation. [Online]. Available from: <http://lltf.net/aptitude-tests/language-aptitude-tests/pimsleur-language-aptitude-battery/> (Accessed January 15, 2017).

Pintrich, P. and De Groot, E. (1990) Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology,* 82 (1), pp. 33-40.

Polio, C. (2012) The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing,* 21 (4), pp. 375-389.

Pollitt, A. (n.d.) *The Oxford Online Placement Test: The meaning of OOPT scores.* Oxford University Press. [Online]. Available from: <https://www.oxfordenglishtesting.com/uploadedFiles/Buy_tests/oopt_meaning.pdf> (Accessed January 22, 2015).

Prat-Sala, M. and Redford, P. (2010) The interplay between motivation, self-efficacy, and approaches to studying. *British Journal of Educational Psychology,* 80 (2), pp. 283-305.

Prat-Sala, M. and Redford, P. (2012) Writing essays: Does self-efficacy matter? The relationship between self-efficacy in reading and in writing and undergraduate students' performance in essay writing. *Educational psychology,* 32 (1), pp. 9-20.

Price, M., Handley, K., Millar, J. and O'Donovan, B. (2010) Feedback : all that effort, but what is the effect? *Assessment & Evaluation in Higher Education,* 35 (3), pp. 277-289.

Puntambekar, S. and Hubscher, R. (2005) Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist,* 40 (1), pp. 1-12.

Purpura, J. (n.d.) *The Oxford Online Placement Test: What does it measure and how?* Oxford University Press. [Online]. Available from: <https://www.oxfordenglishtesting.com/uploadedfiles/6_New_Look_and_Feel/Content/oopt_measure.pdf> (Accessed January 22, 2015).

Qi, D.S. and Lapkin, S. (2001) Exploring the role of noticing in a three-stage second language writing task. *Journal of Second Language Writing,* 10 (4), pp. 277-303.

Qu, S. and Dumay, J. (2011) The qualitative research interview. *Qualitative Research in Accounting and Management,* 8 (3), pp. 238-264.

Rahimi, M. and Zhang, L.J. (2015) Exploring non-native English-speaking teachers' cognitions about corrective feedback in teaching English oral communication. *System,* 55, pp. 111-122.

Rao, Z. and Liu, F. (2011) Effect of academic major on students' use of language learning strategies: a diary study in a Chinese context. *The Language Learning Journal,* 39 (1), pp. 43-55.

Reichelt, M. (2009) Situated writing practices in foreign language settings: The role of previous experience and instruction. In *Writing in Foreign Language Contexts*, ed. R. Manchón, Bristol: Multilingual Matters, pp. 183-206.

Ringbom, H. (1987) *The Role of the First Language in Foreign Language Learning.* Philadelphia: Multilingual Matters Limited.

Rinnert, C.and Kobayashi, H. (2009) Situated writing practices in foreign language settings: The role of previous experience and instruction. In *Writing in Foreign Language Contexts*, ed. R. Manchón, Bristol: Multilingual Matters, pp. 23-48.

Ro, J. (2017) A journey from the classroom to the world of educational reform: A study of three Korean teachers' practitioner inquiry. *Asia Pacific Journal of Education,* 37 (1), pp. 28-41.

Robinson, P. (1995) Attention, memory, and the "Noticing" hypothesis. *Language Learning,* 45 (2), pp. 283-331.

Robinson, P. (1997) Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. *Studies in Second Language Acquisition,* 19 (2), pp. 223-247.

Robinson, S., Pope, D. and Holyoak, L. (2011) Can we meet their expectations? Experiences and perceptions of feedback in first year undergraduate students. *Assessment & Evaluation in Higher Education,* pp. 1-13.

Robson, C. and McCartan, K. (2016) *Real World Research.* Fourth ed. Chichester: John Wiley & Sons Ltd.

Roehr, K. (2005) *Metalinguistic knowledge in second language learning: An emergentist perspective.* Unpublished dissertation ed. Lancaster: Lancaster University.

Roehr, K. (2008) Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics,* 29 (2), pp. 173-199.

Rose, H. and Galloway, N. (in press) Debating standard language ideology in the classroom: Using the 'Speak Good English Movement' to raise awareness of Global Englishes. *RELC Journal*, pp. 1-8.

Roulston, K. (2011) Interview 'problems' as topics for analysis. *Applied Linguistics,* 32 (1), pp. 77-94.

Rubin, H. and Rubin, I. (2012) *Qualitative Interviewing: The Art of Hearing Data.* Third ed. London: SAGE Publications, Inc.

Ruegg, R. (2016) The effect of assessment of process after receiving teacher feedback. *Assessment & Evaluation in Higher Education,* 41 (2), pp. 199-212.

Rushmer, R. and Davies, H.T.O. (2004) Unlearning in health care. *BMJ Quality & Safety,* 13 (suppl 2), pp. ii10-ii15.

S.C.S. (2013) *Test-taking in South Korea: Point me at the SKY.* Online ed. http://www.economist.com/blogs/banyan/2013/11/test-taking-south-korea: The Economist.

S.S.H.S. (2015) *서울세종고등학교/ Seoul Sejong High School.* Seoul Sejong High School. [Online]. Available from: <http://www.schoolinfo.go.kr/ei/ss/Pneiss_b01_s2.do> (Accessed 18 April, 2015).

S.Sci.H.S. (2015) *서울과학고등학교/ Seoul Science High School.* Seoul Science High School. [Online]. Available from: <http://en.sshs.hs.kr/html/curriculum.php> (Accessed 18 April, 2015).

Sachs, R. and Polio, C. (2007) Learners' uses of two types of written feedback on a L2 writing revision task. *Studies in Second Language Acquisition,* 29 (1), pp. 67-100.

Sadler, D.R. (1989) Formative assessment and the design of instructional systems. *Instructional Science,* 18 (2), pp. 119-144.

Sáfár, A. and Kormos, J. (2008) Revisiting problems with foreign language aptitude. *International Review of Applied Linguistics in Language Teaching,* 46, pp. 113-136.

Saldaña, J. (2016) *The Coding Manual for Qualitative Researchers.* Electronic, Third ed. London: Sage.

SamSung (n.d.) *삼성 채용정보 신입사원.* 삼성그룹 SamSung Group. [Online]. Available from: <http://www.samsung.com/sec/aboutsamsung/careers/recruit3.html> (Accessed 19 April, 2015).

Santos, M., López-Serrano, S. and Manchón, R. (2010) The differential effect of two types of direct written corrective feedback on noticing and uptake: Reformulation vs. error correction. *International Journal of English Studies,* 10 (1), pp. 131-154.

Schmidt, R. (1990) The role of consciousness in second language learning. *Applied Linguistics,* 11 (2), pp. 129-158.

Schmidt, R. (1995) Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In *Attention and Awareness in Foreign Language Learning*, ed. R. Schmidt, Honolulu: University of Hawai'i Press, pp. 1-63.

Schmitt, N., Dörnyei, Z., Adolphs, S. and Durow, V. (2004) Knowledge and acquisition of formulaic sequences. In *Formulaic Sequences: Acquisition, Processing, and Use*, ed. S. Norbert, Philadelphia: John Benjamins, pp. 55-86.

Schreier, M. (2012) *Qualitative Content Analysis in Practice.* Electronic ed. London: Sage.

Seidlhofer, B. (2017) Standard English and the dynamics of ELF variation. In *The Routledge Handbook of English as a Lingua Franca*, eds. J. Jenkins W. Baker and M. Dewey, Electronic ed. Abingdon: Routledge, pp. 85-100.

Selinker, L. (1972) Interlanguage. *IRAL; International Review of Applied Linguistics in Language Teaching,* 10 (3), pp. 209-231.

Seth, M.J. (2002) *Education Fever: Society, politics, and the pursuit of schooling in South Korea.* Honolulu: University of Hawaii Press.

Sheen, Y. (2007) The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly,* 41 (2), pp. 255-283.

Sheen, Y. (2011) *Corrective Feedback, Individual Differences and Second Language Learning.* Springer.

Sheen, Y. and Ellis, R. (2011) Corrective feedback in language teaching. In *Handbook of Research in Second Language Teaching and Learning, Volume II.*, ed. E. Hinkel, Hoboken: Taylor & Francis, pp. 593-610.

Sheen, Y., Wright, D. and Moldawa, A. (2009) Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners*. System,* 37 (4), 556-569.

Shepard, L.A. (1995) *Implications for standard setting of the national academy of education evaluation of national assessment of educational progress achievement levels proceedings from the joint conference on standard setting for large-scale assessments.* Washington: National Assessment Governing Board and National Center for Education Statistics.

Shintani, N., Ellis, R. and Suzuki, W. (2014) Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning,* 64 (1), pp. 103-131.

Shintani, N. and Ellis, R. (2015) Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing? *System,* 49, pp. 110-119.

Simard, D., French, L. and Fortier, V. (2007) Elicited metalinguistic reflection and second language learning: Is there a link? *System,* 35 (4), pp. 509-522

Sinclair, J. and Coulthard, M. (1975) *Towards an Analysis of Discourse: The English Used by Teachers and Pupils.* London: Oxford University Press.

Skehan, P. (1998) *A Cognitive Approach to Language Learning.* Oxford: Oxford University Press.

SKKU (n.d.) *학사제도 / Academics System.* 성균관대학교/ SungKyunKwan University. [Online]. Available from: <http://home.skku.edu/~hakbu/haksa/haksa_3_4.htm> (Accessed 19 April, 2015).

SMOE (2011) *Seoul Metropolitan Office of Education.* Available from: <http://english.sen.go.kr/index.jsp> (Accessed July 26, 2013).

SNU (n.d.) *서울대학교: 교육.* Seoul National University. [Online]. Available from: <http://www.snu.ac.kr/graduation-requirement/undergraduate> (Accessed 16 April, 2015).

Song, J. (2011) English as an official language in South Korea: Global English or social malady? *Language Problems & Language Planning,* 35 (1), pp. 33-55.

Spada, N. (1997) Form-focussed instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching,* 30 (2), pp. 73-87.

Spada, N. (2011) Beyond form-focused instruction: Reflections on past, present and future research. *Language Teaching,* 44 (2), pp. 225-236.

Spada, N. and Lightbown, P. (2008) Form-focused instruction: Isolated or integrated? *TESOL Quarterly,* 42 (2), pp. 181-207.

Spada, N., Jessop, L., Tomita, Y., Suzuki, W. and Valeo, A. (2014) Isolated and Integrated form-focused instruction: Effects on different types of L2 knowledge. *Language Teaching Research,* 18 (4), pp. 453-473.

Spada, N. and Tomita, Y. (2010) Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning,* 60 (2), pp. 263-308.

Spencer-Oatey, H. (2000) Introduction: Language, culture and rapport management. In *Culturally Speaking: Managing rapport through talk across cultures.*, ed. H. Spencer-Oatey, London: Continuum, pp. 1-10.

Stajkovic, A.D. and Sommer, S.M. (2000) Self-efficacy and causal attributions: Direct and reciprocal links. *Journal of Applied Social Psychology,* 30 (4), pp. 707-737.

Stevenson, H.W., Chen, C. and Lee, S. (1993) Mathematics achievement of Chinese, Japanese, and American children: Ten years later. *Science,* 259 (5091), pp. 53-58.

Stone, A. (1998) Should we salvage the scaffolding metaphor? *Journal of Learning Disabilities,* 31 (4), pp. 409-413.

Storch, N. (2008) Metatalk in a pair work activity: Level of engagement and implications for language development. *Language Awareness,* 17 (2), pp. 95-114.

Storch, N. (2010) Critical feedback on written corrective feedback research. *International Journal of English Studies,* 10 (2), pp. 29-47.

Storch, N. and Wigglesworth, G. (2010) Learners' processing, uptake, and retention of corrective feedback on writing. *Studies in Second Language Acquisition,* 32, pp. 303-334.

Strother, J. (2012) *Drive for education drives South Korean families into the red.* The Christian Science Monitor. [Online]. Available from: <http://www.csmonitor.com/World/Asia-Pacific/2012/1110/Drive-for-education-drives-South-Korean-families-into-the-red> (Accessed 16 April, 2015).

Suzuki, W. (2012) Written languaging, direct correction, and second language writing revision. *Language Learning,* 62 (4), pp. 1110-1133.

Suzuki, Y. and DeKeyser, R. (2017) The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning,* 67 (4), pp. 747-790.

Swain, M. and Lapkin, S. (1995) Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics,* 16 (3), pp. 371-391.

Swain, M. (1995) Three functions of output in second language learning. In *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson.*, eds. G. Cook and B. Seidlhofer, Oxford University Press, USA, pp. 125-144.

Swain, M. (1998) Focus on form through conscious reflection. In *Focus on Form in Classroom Second Language Acquisition.*, eds. C. Doughty and J. Williams, Cambridge: Cambridge University Press, pp. 64-81.

Swain, M. (2006) Languaging, agency and collaboration in advanced second language proficiency. In *Advanced Language Learning: The Contribution of Halliday and Vygotsky.*, ed. H. Byrnes, London: Continuum International Publishing, pp. 95-108.

Swain, M. and Lapkin, S. (1998) Interaction and second language learning: Two adolescent French immersion students working together. *The Modern Language Journal,* 82 (3), pp. 320-337.

Swain, M. and Lapkin, S. (2002) Talking it through: two French immersion learners' response to reformulation*. International Journal of Educational Research*, 37 (3-4), pp. 285-304

Swain, M., Lapkin, S., Knouzi, I., Suzuki, W. and Brooks, L. (2009) Languaging: University students learn the grammatical concept of voice in French. *The Modern Language Journal,* 93 (1), pp. 5-29.

Talisse, R. and Aikin, S. (2011) Introduction. In *The Pragmatism Reader*, eds. R. Talisse and S. Aikin, Princeton: Princeton University Press, pp. 1-11.

Taras, M. (2001) The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education,* 26 (6), pp. 605-614.

Thompson, I. (2009) Scaffolding in the writing center: A microanalysis of an experienced tutor's verbal and nonverbal tutoring strategies. *Written Communication,* 26 (4), pp. 417-453.

Thornbury, S. (1997) Reformulation and reconstruction: Tasks that promote 'noticing'. *ELT Journal,* 51 (4), pp. 326-335.

Thorne, S. (2005) Epistemology, politics, and ethics in sociocultural theory. *The Modern Language Journal,* 89 (3), pp. 393-409.

Tian, M. and Lowe, J. (2013) The role of feedback in cross-cultural learning: a case study of Chinese taught postgraduate students in a UK university. *Assessment & Evaluation in Higher Education,* 38 (5), pp. 580-598.

Toh, P.C., Leong, Y.H., Toh, T.L., Dindyal, J., Quek, K.S., Tay, E.G. and Ho, F.H. (2014) The problem-solving approach in the teaching of number theory. *International Journal of Mathematical Education in Science and Technology,* 45 (2), pp. 241-255.

Tomlin, R.S. and Villa, V. (1994) Attention in cognitive science and second language acquisition. *Studies in Second Language Acquisition,* 16 (2), pp. 183-203.

Tran, T.T. (2013) Is the learning approach of students from the Confucian heritage culture problematic? *Educational Research for Policy and Practice,* 12 (1), pp. 57-65.

Truscott, J. (1996) The case against grammar correction in L2 writing classes. *Language Learning,* 46 (2), pp. 327-369.

Truscott, J. and Hsu, A.Y. (2008) Error correction, revision, and learning. *Journal of Second Language Writing,* 17 (4), pp. 292 305.

Tsui, A.B. (1985) Analyzing input and interaction in second language classrooms. *RELC Journal,* 16 (1), pp. 8-30.

Ur, P. (1996) *A Course in Language Teaching: Practice and Theory.* Cambridge: Cambridge University Press.

Van Beuningen, C., De Jong, N.H. and Kuiken, F. (2012) Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning,* 62 (1), pp. 1-41.

VanPatten, B. (2016) Why explicit knowledge cannot become implicit knowledge. *Foreign Language Annals,* 49 (4), pp. 650-657.

VanPatten, B. (2002) Processing instruction: An update. *Language Learning,* 52 (4), pp. 755-803.

Vardi, I. (2009) The relationship between feedback and change in tertiary student writing in the disciplines. *International Journal on Teaching and Learning in Higher Education,* 20 (3), pp. 350-361.

Vaughan, E. (2012) Discourse analysis. In *Research Methods and Methodologies in Education.*, eds. J. Arthur M. Waring R. Coe and L. Hedges, London: Sage, pp. 272-282.

Vélez-Rendón, G. (2002) Second language teacher education: A review of the literature. *Foreign Language Annals,* 35 (4), pp. 457-467.

Vogt, W., Gardner, D. and Haeffele, L. (2012) *When to Use What Research Design.* London: The Guilford Press.

Wang, C. and Wang, M. (2015) Effect of alignment on L2 written production. *Applied Linguistics*, 36 (5), pp. 503-526.

Wang, W. and Loewen, S. (2016) Nonverbal behavior and corrective feedback in nine ESL university-level classrooms. *Language Teaching Research,* 20 (4), pp. 459-478.

Wason, P. (1960) On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology,* 12 (3), pp. 129-140.

White, L., Spada, N., Lightbown, P. and Ranta, L. (1991) Input enhancement and L2 question formation. *Applied Linguistics,* 12 (4), pp. 416-432.

Wigglesworth, G. and Storch, N. (2012) What role for collaboration in writing and writing feedback. *Journal of Second Language Writing,* 21 (4), pp. 364-374.

Williams, J. (2004) Tutoring and revision: Second language writers in the writing center. *Journal of Second Language Writing,* 13 (3), pp. 173-201.

Williams, C. (2017) *Teaching English in East Asia: A teacher's guide to Chinese, Japanese, and Korean learners.* Singapore: Springer.

Williams, J. and Kane, D. (2009) Assessment and feedback: Institutional experiences of student feedback, 1996 to 2007. *Higher Education Quarterly,* 63 (3), pp. 264-286.

Winstone, N.E., Nash, R.A., Rowntree, J. and Parker, M. (2017) 'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and recipience. *Studies in Higher Education, 42 (11),* pp. 2026-2041.

Winters, S. and O'Brien, M.G. (2015) *Personal communication.*

Winters, S. and O'Brien, M.G. (2013) Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Communication,* 55 (3), pp. 486-507.

WISET (2015) *2015 Report on Korean Women in Science, Engineering and Technology.* Seoul - http://www.eng.wiset.or.kr: WISET: Women in Science, Engineering and Technology.

Wong, J. (2004) Are the learning styles of Asian international students culturally or contextually based? *International Education Journal,* 4 (4), pp. 154-166.

Wong, L.L.C. and Nunan, D. (2011) The learning styles and strategies of effective language learners. *System,* 39 (2), pp. 144-163.

Wood, D., Bruner, J.S. and Ross, G. (1976) The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry,* 17 (2), pp. 89-100.

Wood, R.E. and Locke, E.A. (1987) The relation of self-efficacy and grade goals to academic performance. *Educational and Psychological Measurement,* 47 (4), pp. 1013-1024.

Woods, D. (2006) The social construction of beliefs in the language classroom. In *Beliefs About SLA: New research approaches*, eds. P. Kalaja and A.M.F. Barcelos, New York: Springer, pp. 201-229.

Robert Woore. (2011) *Investigating and developing beginner learners' decoding proficiency in second language French: An evaluation of two programmes of instruction.* Unpublished doctoral thesis. University of Oxford.

Xu, C. (2009) Overgeneralization from a narrow focus: A response to Ellis et al. (2008) and Bitchener (2008). *Journal of Second Language Writing,* 18 (4), pp. 270-275.

Yoshida, R. (2010) How do teachers and learners perceive corrective feedback in the Japanese language classroom? *The Modern Language Journal,* 94 (2), pp. 293-314.

Yu, A. (1996) Ultimate life concerns, self, and Chinese achievement motivation. In *The Handbook of Chinese Psychology*, ed. M. Bond, Hong Kong: Oxford University Press, pp. 227-246.

Zawacki, T.M.and Habib, A.S. (2014) Negotiating "errors" in L2 writing: Faculty dispositions and language difference. In *WAC and Second Language Writers: Research Towards Linguistically and Culturally Inclusive Programs and Practices*, eds. T.M. Zawacki and M. Cox, electronic ed. The WAC Clearinghouse: Parlor Press and the WAC Clearinghouse, pp. 183-210.

Zimbardi, K., Colthorpe, K., Dekker, A., Engstrom, C., Bugarcic, A., Worthy, P., Victor, R., Chunduri, P., Lluka, L. and Long, P. (2017) Are they using my feedback? The extent of students' feedback use has a large impact on subsequent academic performance. *Assessment & Evaluation in Higher Education*, 42 (4), pp. 625-644.

# Appendices

Table of Contents for appendices


| | | |
|---|---|---|
| Appendix A | Participant writing sample | 272 |
| Appendix B | Self-efficacy measure | 273 |
| Appendix C | Language aptitude test | 274 |
| Appendix D | Reflective worksheet | 276 |
| Appendix E | Interview protocol | 277 |
| Appendix F | Background survey | 278 |
| Appendix G | Oxford Online Proficiency Test example | 280 |
| Appendix H | Grammar correction pre-test | 282 |
| Appendix I | Conference transcript | 283 |
| Appendix J | Tsui's (1985) Seventeen-Category System | 284 |

The appendices originally presented here cannot be made freely available via ORA because of copyright. The copyrighted items were sourced from the authors as referenced in the text starting on p. 100 of the thesis.