

THE INTERACTION BETWEEN GRAMMATICAL KNOWLEDGE AND EXPLICITNESS  
IN L2 WRITTEN CORRECTIVE FEEDBACK

By Dan Brown

A Dissertation

Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in Applied Linguistics

Northern Arizona University

May 2017

Approved:

William Crawford, Ph.D., Chair

Luke Plonsky, Ph.D.

Soo Jung Youn, Ph.D.

Yuly Asención Delaney, Ph.D.

## ABSTRACT

### THE INTERACTION BETWEEN GRAMMATICAL KNOWLEDGE AND EXPLICITNESS IN L2 WRITTEN CORRECTIVE FEEDBACK

DAN BROWN

This study investigates the relative effectiveness of different types of written corrective feedback (WCF) relative to error type and grammatical knowledge in developing second language (L2) accuracy in writing. Findings on the effectiveness of WCF have been mixed regarding which types of feedback are more effective, however, recent evidence has suggested potential benefits for direct methods (i.e., supplying students with the correct forms) in developing grammatical accuracy over time (Kang & Han, 2015; Russell & Spada, 2006; van Beuningen, de Jong, & Kuiken, 2012). The opinions of several researchers, teachers, and students, however, support the use of indirect WCF (i.e., indicating the location and type of error for students through use of a code) because it is argued to encourage students' analytic reflection, engagement, and processing of the feedback they receive (Ferris, 2010, 2011; Lalande, 1982). Several learner-internal and learner-external variables have been hypothesized to moderate the effectiveness of different WCF methods (e.g., learners' perceptions, proficiency, developmental readiness, the nature of the linguistic target, grammatical knowledge), with limited evidence to support any of these claims. This study reports on a controlled classroom-based quasi-experiment conducted in an EFL university context ( $N = 127$ ) that investigates the role of grammatical knowledge in moderating the effectiveness of direct vs. indirect WCF on six target grammatical error types. Results provide further evidence in support of moderate effectiveness of WCF for both direct (Cohen's  $d = .28$ ) and indirect methods (Cohen's  $d = .43$ ), which outperformed the control after five treatments. No clear relationship was found between prior

grammatical knowledge and gains in accuracy at the group level, although further analysis of the top and bottom quartile of test scorers across error types suggests advantages for direct WCF for students with prior grammatical knowledge. Pedagogical and theoretical implications are discussed, as well as suggestions for future research on WCF.

***Key words:*** written corrective feedback, feedback type, linguistic targets, grammatical knowledge, metalinguistic knowledge

Copyright by  
Dan Brown  
2017



## ACKNOWLEDGEMENTS

This project has benefited from the help of many, to whom I would like to extend my gratitude. Dr. William Crawford, chair of my advisory committee, has been unwavering in his support throughout this study. His encouragement and generosity with his time from the beginning of the project have set an example for how I hope to mentor students. His support in the planning of the project, skill in designing classroom research and in analyzing data were invaluable. Dr. Luke Plonsky has been an inspiration and mentor since I began my PhD studies, and provided invaluable advice for this study. I have greatly benefited from his expertise and leadership in methodological concerns and knowledge in statistics, not only for this study, but in forming my general research interests. Dr. Soo Jung Youn has provided great insight in helping to guide the design of the assessment as well many of the methodological decisions, and her support has been crucial when I needed it most. Dr. Yuly Asención Delaney has provided generous and insightful feedback to help improve the study with each draft. Her experience in assessment and feedback in second language writing, as well as her attention to detail of the manuscript has strengthened the design of the study, and I am grateful for all her help.

The data collection for this study could not have occurred without the support and help of Dr. Alexander Nanni. His support for this project was indispensable, not only in paving the way for me to join his program as a visiting instructor for the summer of 2016, but also in helping to participate with data collection in his writing class and ensuring the success of the data collection at several junctures. His care, attention to details, and knowledge of research design helped the validity of the methods in this study. He also provided me with invaluable professional development opportunities by participating in his program in ways that made my time in Thailand incredibly valuable beyond my dissertation data collection. I am also grateful for the

generous help of Douglas Black and Adriano Quieti for their participation by helping with data collection. Their unwavering patience and willingness to accommodate made it possible to carry out this study.

I would also like to thank those who assisted in the coding of the data. Michael Suhan was a tremendous help when I was in desperate need by providing much of the feedback treatment with careful attention to detail, and generously shared his expertise in computer programming to develop a program to automatically tally error types in the coding, which improved accuracy. My thanks also go to Maren Greve, who helped as a research assistant as a second coder in scoring metalinguistic knowledge responses and helped to improve the rubric.

I also thank my mother for encouraging each step of my long and winding path, and finally, my wife Ingrid, who has helped in organizing thousands of writing files and has shown never ending patience with the schedule of a PhD student conducting a dissertation in our first year of marriage. She has supported my every move with kindness and love.

## TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGMENTS.....	v
TABLE OF CONTENTS.....	vii
LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi
CHAPTER 1 INTRODUCTION.....	1
CHAPTER 2 REVIEW OF THE LITERATURE.....	6
Overview of Research on the Effectiveness of Written Corrective Feedback in the Development of Accuracy.....	6
Explicitness in WCF Strategies: The Effects of Direct vs Indirect Feedback Types.....	8
Variables Associated with the Utilization of WCF.....	10
Linguistic Target Relative to WCF Type.....	12
Developmental Readiness.....	17
Aptitude-Treatment Interaction Research.....	18
The Potential Role of Grammatical Knowledge on the Efficacy of WCF Types.....	19
Defining and Measuring Grammatical Knowledge.....	22
Targeted error types.....	23
Conceptualizing grammatical knowledge.....	24
Grammatical knowledge in the context of WCF.....	25
Methods of assessing grammatical knowledge.....	25
Defining and operationalizing metalinguistic knowledge.....	26
Scoring criteria for metalinguistic knowledge.....	30
Disparity between error types.....	32
Summary of the Literature Review .....	34
Research Questions.....	35
CHAPTER 3 METHODOLOGY.....	36
General Research Design.....	36
Variables in the Study.....	36
Research Setting.....	37
Participants and Grouping.....	39
Writing Tasks.....	41
Selection and Rationale for the Target Error Categories.....	43
Comprehensive without overwhelming in scope.....	43
Rule-governed structures.....	44
Representative of different categories of structures.....	44
Most prominent for the target population.....	44
Representative of previous research.....	46



Treatment.....	47
Inter-Coder Reliability for Feedback Provision.....	49
Grammatical Knowledge Test Development.....	50
Test format.....	50
A bottom-up approach to item validation.....	51
Piloting of the test.....	59
Data Collection Procedures.....	59
Test administration.....	60
Training students to use indirect coded feedback.....	60
Writing data collection.....	62
Analysis.....	63
Linguistic measures of the dependent variable.....	63
Statistical analyses.....	64
 CHAPTER 4 RESULTS.....	 67
Comparability of Groups at the Onset of the Study.....	67
Results of Research Question 1: Effectiveness of WCF Types.....	70
Results of Research Question 2: Interaction between WCF Type and Error Type.....	73
Results of Research Question 3: The Relationship between Grammatical Knowledge and Effectiveness of WCF.....	77
Results of Research Question 4: Interaction between Grammatical Knowledge, Feedback Type, and Development in Grammatical Accuracy.....	84
Post-hoc analysis.....	88
Summary of Results.....	92
 CHAPTER 5 DISCUSSION.....	 94
Research Question 1: Relative Effectiveness of WCF Types.....	94
Research Question 2: Interaction between WCF Type and Error Type.....	99
Research Questions 3 and 4: The Relationship between Grammatical Knowledge and Effectiveness of WCF.....	104
 CHAPTER 6 CONCLUSION.....	 109
Theoretical Implications.....	109
Pedagogical Implications.....	109
Limitations of this Study.....	112
Suggestions for Future Research.....	113
 REFERENCES.....	 116
 APPENDICES.....	 128
Appendix A. Grammatical Knowledge Test and Answer Key .....	128
Appendix B. Table of Specifications for Grammatical Knowledge Test.....	138
Appendix C. Training Materials for Coded Feedback.....	139
Appendix D. Grammar Key for Students (Indirect Coded Feedback Group).....	141
Appendix E. IRB Permission.....	142

## LIST OF TABLES

Table	Page
1	Studies Investigating Truscott’s (2007) Error Type “Treatability” Claim.....15
2	Error Types Investigated in Previous WCF Studies that Include Comprehensive Feedback.....23
3	Instruments Designed to Measure Explicit L2 Grammatical Knowledge.....27
4	Metalinguistic Knowledge Test Designs.....29
5	Example Scoring Scale for the MK Rule Explanation.....31
6	Examples of Scoring Criteria for Specific Error Categories.....31
7	Description of Study Variables.....37
8	Essay Topics.....41
9	Description of Error Categories and Codes Used for Feedback and Analysis.....43
10	Most Frequent Error Types in Level 3 Writings.....45
11	Common Error Patterns in Previous Level 3 Student Essays.....53
12	Test Items (organized by item type within error categories).....56
13	Equivalency between Groups for Pretest Accuracy.....68
14	Descriptive Statistics of Accuracy by Individual Essay.....69
15	Comparison of Overall Gains in Accuracy between Groups.....72
16	Gains in Accuracy between Pre- and Posttests.....74
17	Descriptive Statistics for the Grammatical Knowledge Test Results.....79
18	Grammatical Knowledge Test Item Analysis.....79
19	Reliability for Grammatical Knowledge Test by Test Section.....81
20	Inter-rater Reliability for Metalinguistic Knowledge Scoring.....82

21	Grammatical Knowledge Test Results (by error type).....	84
22	Metalinguistic Knowledge Section Test Results (by error type) .....	85
23	Correlation Matrix (Grammatical Knowledge x Gains in Accuracy).....	86
24	Correlation Matrix (Metalinguistic Knowledge x Gains in Accuracy).....	87
25	Post-hoc Analysis: Gains in Accuracy for Highest/Lowest Grammatical Knowledge Groups.....	89

## LIST OF FIGURES

Figure		Page
1	Example of Indirect Feedback Treatment.....	48
2	Data Collection Procedures.....	62
3	Overall Gains in Accuracy by Treatment Type.....	72
4	Group Gains in Accuracy by Error Type.....	75
5	Changes in Overall Accuracy Comparing Highest and Lowest Overall Test Performers that Received Direct Feedback.....	90
6	Changes in Accuracy for Articles Comparing Highest and Lowest Test Performers in the Direct Feedback Group.....	91

## CHAPTER 1

### INTRODUCTION

Framed within a cognitive perspective of second language acquisition (SLA), this study is motivated by two themes in SLA research: (a) the repeated calls for investigating the variables constraining the effectiveness of corrective feedback (e.g., Ellis & Sheen, 2006; Ferris, 2004; Gu nette, 2007; Russel & Spada, 2006; Storch, 2010) and (b) the discord in research findings related to effectiveness of written corrective feedback (WCF) (Kang & Han, 2015; Liu & Brown, 2015). From a practical perspective, although accuracy is not a pressing goal for many second language (L2) learners, those preparing for the demands of academic and professional writing have genuine needs to improve their written accuracy, as it can stigmatize L2 writers even when there is no obstruction to comprehension. Further, most L2 learners expect their writing teachers to provide WCF (Ferris, 2011), as it may be one the only opportunities for L2 learners to focus on certain features of language form in their language study. Therefore, WCF remains a central concern in L2 writing classrooms.

From a theoretical perspective, it is generally accepted that adult L2 learning requires some degree of negative input in the form of corrective feedback to guide learners to notice what is lacking in their production and modify their interlanguage (Gass & Varonis, 1994; Schmidt, 2001), possibly as a result of established explicit knowledge (DeKeyser, 2007). Evidence is more developed for the role of corrective feedback in oral production (Plonsky & Brown, 2015), while in L2 writing the debate has remained contentious for many years. Only recently have meta-analytic findings (e.g., Kang & Han, 2015) pointed to moderate effect for WCF in improving learners' development in accuracy over time, yet despite numerous published studies

investigating WCF, controversy persists as to which approaches best help student writers sustain their development in accuracy across a range of linguistic features.

Since Truscott (1996) intensified the debate over the value of WCF by claiming that it is not only ineffective but potentially counterproductive for L2 students, researchers have pushed to advance methodological design (e.g., Ferris, 2004; Gu nette, 2007; Liu & Brown, 2015; Storch, 2010) and have focused attention on the variables that could influence how and when WCF can be most effective. As Goldstein (2004) illustrates, a combination of factors play a role that can be found at the institutional or program level, in writing task requirements, and that teachers and students bring to the process. Perhaps the most commonly studied variable in this line of research is feedback type, typically dichotomized between direct (i.e., supplying students with the correct forms) and indirect methods (i.e., indicating the location and/or type of error for students, often using a code). Findings have been mixed, though a few recent studies have reported indistinguishable benefits for direct and indirect WCF in the short-term (for revision) and advantages for direct WCF for long-term development in grammatical accuracy in new writings (Bitchener & Knoch, 2010; van Beuningen, de Jong, Kuiken, 2012). The advent of meta-analytic research in applied linguistics has helped clarify evidence that supports the overall effectiveness of WCF in promoting accuracy in students' writing (Biber, Nekrasova, & Horn, 2011; Kang & Han, 2015; Russell & Spada, 2006), and these results further suggest an advantage for direct methods over indirect methods for long-term development. However, several researchers and teachers support the use of indirect WCF because it is thought to encourage students' analytic reflection, engagement, and processing of the feedback they receive (Ferris, 2010, 2011; Lalande, 1982). With few studies that have compared individual linguistic targets and other potential moderators of effectiveness, is difficult to explain divergent findings.

Other variables hypothesized to moderate the relative effectiveness of different feedback methods include proficiency (Park, Song & Shin, 2016; van Beuningen et al., 2012), previous language instruction (Ferris & Hedgcock, 2004), pedagogical focus (Russell & Spada, 2006), motivation (Hyland & Hyland, 2006; Storch, 2010) and the nature of the targeted error types (Ferris, 1999; Truscott, 2007), hypothesizing that some linguistic structures may not be amenable to corrective feedback, and that some are more conducive to rule-based explanation and therefore may benefit from indirect methods (Ferris, 1999). Limited empirical research has explored these variables to clarify the relative effectiveness of different feedback types (see van Beuningen et al., 2012 for an exception).

There are only a few qualitative studies that have focused on how learners engage with the feedback they receive, which have found students to struggle with utilizing indirect feedback because they are unfamiliar with the grammatical rules and terminology connected with various error categories that teachers select, particularly when a wide scope of categories are targeted (Ferris, 1995; Lee, 1997). Some evidence (Ferris, 2006) also suggests that teachers employ a mixed approach of both direct and indirect feedback relying on intuitive judgment as to which WCF type will benefit students in each instance. In follow-up interviews with teachers after a semester of feedback, Ferris (2006) discovered that teachers indicated their awareness (consciously or unconsciously) of students' ability to deal with the feedback they receive, and how this awareness influenced the amount and type of feedback they provided. Her analysis revealed that teachers were more likely to supply indirect feedback to "treatable" errors (defined as errors that are rule-governed, such as verb tense/form, subject-verb agreement, run-ons, fragments, noun endings, articles, and pronouns) and direct feedback to "untreatable" errors (i.e., more idiomatic error categories that do not follow straightforward grammar rules, such as word-

choice and sentence structure), suggesting that teachers may intuitively supply direct corrections for errors they sense that students are less likely to be able to self-correct. One of the few studies to investigate the notion of “treatability” in the development of accuracy in new writings was van Beuningen et al. (2012), which found that non-grammatical error types (operationalized as word choice, spelling, punctuation, and capitalization) benefited most from indirect WCF, while grammatical errors (inflectional errors, articles, word order errors, omissions of necessary elements, fragments, additions of non-necessary elements, pronominal errors) resulted in durable accuracy gains in response to direct WCF.

Rather than simply label error types as more or less treatable for all learners, one goal of the present study is to problematize these findings by considering whether preexisting grammatical knowledge plays a role in predicting treatability of certain errors according to feedback type. Although a few recent studies have investigated the interaction between individual learner variables (e.g., working memory, language learning aptitude) and different oral feedback types (e.g., Li, 2013; Yilmaz, 2013), no research to date has measured the grammatical knowledge that students bring to a writing classroom to better understand how such knowledge may interact with their ability to utilize different types of feedback to develop their accuracy over time. This study hypothesizes that learners will benefit more from indirect feedback on error types for which they have more robust grammatical understanding; and conversely, that they will benefit more from direct feedback on error types for which they lack the requisite grammatical knowledge to be able to utilize less explicit feedback. In other words, indirect feedback may be more effective in helping students develop grammatical accuracy when sufficient grammatical knowledge is present (i.e., their ability to recognize the meaning of an indirect code and to understand and operationalize grammatical rules to make corrections) that



can allow learners to self-repair, while direct methods may be more effective when learners lack the grammatical knowledge to make use of coded feedback. Findings could inform practitioners as to how and when different WCF strategies can help L2 student writers. If teachers could use a diagnostic instrument, such as the test presented in this study, to quickly assess students' relative grammatical knowledge among common error categories, feedback could be more effectively tailored to encourage development in accuracy over the course of a semester. And if a relationship is found between grammatical knowledge and development in accuracy, it would support the role of explicit instruction as a complement (or prerequisite) to feedback on grammatical accuracy for certain grammatical structures.

## CHAPTER 2

### REVIEW OF THE LITERATURE

#### **Overview of Research on the Effectiveness of WCF in the Development of Accuracy**

One of the most contentious debates in L2 writing research continues over the value of WCF in L2 writing pedagogy. While providing feedback on grammatical accuracy remains a central role for many L2 writing teachers, surprisingly limited consensus has been reached on fundamental issues of how WCF can most effectively be applied in L2 classrooms to help learners improve their written accuracy, particularly in terms of the most appropriate strategies to match instructional contexts. Research questions investigating WCF have focused on whether it can improve revision and/or lead to development in accuracy in the long-term in new writings, the circumstances and situations constraining levels of effectiveness, pedagogical issues (e.g., which types of WCF are most effective, which errors should be corrected, when and how feedback should be provided), and teacher and learner perspectives on its value (Bitchener & Ferris, 2012). Debate persists as methodological reviews have pointed to the challenges and inconsistencies in the this domain highlighting issues such as a common lack of control group, exclusive focus on revision rather than development in new pieces of writing, prevalence of single-shot treatments, and incomparability between studies due to inconsistent treatments and accuracy measures (Bruton, 2009; Ellis, Sheen, Murakami, & Takashima, 2008; Ferris, 2004; Guénette, 2007, Liu & Brown, 2015; Storch, 2010; Truscott, 2007; van Beuningen, 2010).

The last decade has seen several studies with more robust experimental design that sought to address the shortcomings of earlier studies (e.g., Bitchener, 2008; Bitchener & Knock, 2008, 2010; Chandler, 2003; Ellis et al., 2008; Sheen, 2007, 2010; Sheen et al., 2009), particularly the need for a control group which Liu and Brown (2015) have shown lacking in

over a third of studies that investigate future gains in accuracy. While these studies have improved many of the shortcomings of earlier work, most limited the scope of targeted error categories to only one or a few types. Two exceptions are Ellis et al. (2008) and Sheen et al. (2009), which included the comparison of unfocused and focused feedback within a single study design. In general, the wave of studies through the 2000's provided more consistent results in support of the benefits of WCF in long-term development. Debate continued, however, because many of these studies focused on a single error type, which intuitively is more likely to reveal greater gains with less cognitive burden on learners and increased opportunity for focused attention (Bruton, 2009; Ferris, 2010), processes long argued supportive of second language acquisition (Schmidt, 2001). While these SLA-oriented studies that investigate a single error type (usually English article usage) have shown focused feedback to be more sensitive to long-term gains in accuracy, their classroom ecological validity has been called to question (Bruton, 2009; Ferris, 2010; Liu & Brown, 2015; van Beuningen et al., 2012), as students' aim to develop their overall accuracy in writing classes, not only for isolated error categories. The fact that these studies focused only on WCF within a single linguistic environment limits their generalizability to other linguistic domains (Xu, 2009).

A more recent study that represents a response to the criticisms in research design and that investigates comprehensive feedback on a wide range of error types is van Beuningen et al. (2012). They included two contrast groups (both to compare no feedback and writing practice alone) and used a pre-, post-, and delayed posttest design to investigate long-term development in the L2 learning of Dutch at the secondary level and found significantly fewer errors in both revised and new texts for both direct and indirect forms of feedback. Further support for the benefits of WCF continues to build with the advent of research synthesis in applied linguistics

(e.g., Norris & Ortega, 2000; 2006). Meta-analytic studies have aggregated findings in this area to reveal substantial overall benefits of WCF (Biber et al., 2011; Kang & Han, 2015; Russell & Spada, 2006). Most recently, Kang and Han (2015) meta-analyzed the results of 22 published studies that specifically investigated the effectiveness of WCF in developing grammatical accuracy in new writings finding an overall moderate to large effect (*Hedge's*  $g = .68$ ,  $SE = .13$ ,  $CI = .42-.93$ ,  $p < .0001$ ). These findings refute claims by Truscott over the years (1996, 2007) that WCF is ineffective in supporting long-term development in new texts, and begin to shed light on moderator variables that support or limit the effectiveness of WCF. In addition to the scope of the range of errors under investigation, a central question concerns which type of feedback is most effective. The following section reviews the research on WCF type, typically distinguished between direct and indirect techniques.

### **Explicitness in WCF Strategies: The Differential Effects of Direct vs. Indirect Feedback**

Most WCF studies have compared the effectiveness of different WCF types, which generally focus on the level of explicitness of feedback (i.e., the extent to which teachers provide information within the feedback that helps learners to correct an error). Explicitness of feedback type is generally viewed in terms of the dichotomy between direct and indirect methods, although a wider range of options have been studied within these categories. Although the terms “direct” and “indirect” are not always used consistently in the literature, direct WCF has typically been defined as “that which provides some form of explicit correction of linguistic form or structure above or near the linguistic error and usually involves the crossing out of an unnecessary word/phrase/morpheme, the insertion of a missing word/phrase/morpheme, and/or the provision of the correct form or structure” (Bitchener & Ferris, 2012, p. 65). Indirect WCF, on the other hand, indicates an error has been made but does not provide a correction. Indirect

WCF is operationalized in different ways, such as underlining or circling errors, recording the number of errors in a given line in the margin, or more commonly, through the provision of a code to show the category of error (e.g., “SV” for subject-verb agreement). Although coded feedback can be considered more explicit than simply underlining an error, it is still less explicit than direct correction because rather than being provided the correct form, learners are left to resolve and self-correct the problem that is drawn to their attention.

The limited body of research that has investigated the short and long term effects of these two contrasting methods of WCF has shown conflicting results. One of the earliest studies to compare the effectiveness of the two approaches was Lalande (1982), which reported an advantage for indirect codes that was not statistically significant, however the two treatments differed in more respects than just the method of the feedback. Later, Semke (1984) found no difference between direct and indirect feedback, although the groups in this study received different quantities of writing and only one group had to re-write. Ferris (2006) found that while direct correction led to higher percentage of successful short-term revisions, greater gains were found in long-term written accuracy as a result of indirect feedback, although this study did not include a control group. In contrast, more recent studies with more robust experimental design have pointed to the advantages of direct feedback in development of accuracy. Bitchener and Knock (2010) found direct and indirect options equally effective in the short term, but direct more effective for long-term acquisition. Van Beuningen et al.’s (2008, 2012) studies found equivalent results, generally supporting direct WCF for long-term gains, particularly for structures that are rule-based. Finally, moderator analysis in Kang and Han’s (2015) meta-analysis, which explores whether certain variables are associated with differences in the effects, found no statistical differences between feedback types in their ability to promote accuracy gains

in new writings, but aggregate effect sizes of direct feedback ( $g = .60, k = 12$ ) were higher than indirect ( $g = .36, k = 5$ ), suggesting a potential overall advantage for direct WCF in helping students to improve their accuracy. These results represent the closest answer to the question of relative effectiveness between WCF types; however, the differences between direct and indirect types was not statistically different from a meta-analytic perspective. Practitioners looking for guidance on what type of feedback to provide to their students are likely to be left conflicted, with limited evidence in either direction. For these reasons, it has been pointed out that a mixed approach to feedback type might prove most useful, varying according to a number of factors including error type, proficiency, and linguistic/grammatical knowledge (Chaney, 1999; Ferris, 1999; Hendrickson, 1980). The following section explores the factors that have been investigated that likely moderate the effectiveness of different WCF techniques.

### **Variables Associated with the Utilization of WCF**

It is important for researchers and teachers to better understand the factors that influence the use and effectiveness of WCF so results can be accurately interpreted, particularly in classroom contexts with a several contextual factors at play. Practitioners likewise should be concerned with the contextual variables that promote or limit the benefit that results from all the time that goes into providing written feedback on accuracy. With a better understanding of the contextual factors that influence the utilization of feedback, teachers can decide which factors can be accommodated and how. Ellis (2010), Goldstein (2004) and Li (2010) argue convincingly for a framework of corrective feedback research that accounts for contextual, learner, and teacher variables that can be separated into learner external factors (e.g., program and course requirements regarding accuracy and WCF, scoring rubrics, instructor attitudes, feedback type, quality of feedback, linguistic targets) and learner internal factors (e.g., learner attitudes and

beliefs towards writing, accuracy or feedback types, learning style, memory, motivation, ability to notice and reflect on errors and feedback, aptitude or knowledge required to self-correct), as well as an interaction between multiple factors, such as student engagement with the feedback they receive (Ellis, 2010).

A complex and dynamic interaction between several factors likely influence corrective feedback, although these influences have more commonly been studied in oral CF studies (e.g., Li, 2010; Lyster & Saito, 2010; Lyster, Saito, & Sato, 2013) than WCF studies (see Ferris, 2006, Rahimi, 2014, and Storch & Wigglesworth, 2010, for exceptions). Oral feedback studies have begun to explore the relative effectiveness of different feedback strategies relative to individual differences in aptitude-treatment interaction SLA studies (e.g., Li, 2013; Yilmaz, 2013), but this line of research has yet to be explored in WCF. Most WCF studies have focused exclusively on the effect of feedback types on learning outcomes with limited attention to intervening factors in the complex relationship between learners, feedback, and context (Gu enette, 2007; Liu & Brown, 2015, Storch, 2010).

Attention to individual learner variables has focused most on learner and teacher perceptions, such as preferences for different WCF types. Survey studies have shown student preferences for both indirect feedback (Ferris, Chaney, Komura, Roberts, 2000; Komura, 1999), as well as for direct correction (Ferris, 1995), although not enough of these studies can be compared to generalize understanding of the role of context. Other studies have considered learner motivation as a moderator (e.g., Ferris & Roberts, 2001; Liu, 2016), a confounding variable often overlooked in this line of research. It remains unclear, though, how preference or motivation moderate the effectiveness of different feedback types. The following sections explore research on two other variables thought to influence learners' use of WCF: linguistic target and

developmental readiness, which serve to introduce a central variable in this study—grammatical knowledge.

**Linguistic Target Relative to WCF Type.** The ways that different linguistic targets (i.e., error types) may moderate effectiveness of corrective feedback—in general, and relative to different feedback strategies—has been explored to a limited extent in oral and written feedback research. In interactionist SLA research on oral feedback, variation has been observed in terms of development for different linguistic features and for learners with varying proficiency. Mackey (2006) found variation in noticing of and subsequent development among three linguistic features in response to oral feedback in interaction (questions, singular-plural noun agreement, and past tense verb tense). Feedback on question formation errors were most often noticed and successfully acquired in this study, followed by noun agreement and finally past verb tense. She attributed these differences among error types to perceptual saliency, occurrence frequency, and communicative value. In another oral CF study that explored linguistic target, Li (2010) found interaction between error type, proficiency and efficacy, as less advanced learners benefited more from recasts (note that recasts in oral CF can be considered similarly to direct WCF, although perhaps less salient) for simple structures in learning L2 Chinese (classifiers), but not for more complex structures (perfective markers).

There is surprisingly little research that focuses on the relative amenability of different error types to written corrective feedback. One reason for this is that a good portion of the existing quasi-experimental studies that investigate the development of accuracy over time have focused on a single error type, particularly English article use (e.g., Bitchener & Knoch, 2008; 2010; Ellis et al., 2008; Sheen, 2007; Sheen et al., 2009). Several other studies in this domain have only focused on a few error types and those that cover a wider range of error categories



(i.e., those that have taken an unfocused approach) have also examined a range of treatment types that do not always line up, and often apply inconsistent classification of errors, making it difficult to compare results across research (Liu & Brown, 2015).

Hypotheses and studies of error type as a variable in the effectiveness of WCF highlight the potential differential effects between different domains of linguistic knowledge (i.e., morphological, syntactic, and lexical) or structure (e.g., word, phrasal, clausal) in response to WCF. In fact, when Truscott fueled the debate over the value of WCF beginning in 1996, part of his theoretical argument rested on the claim that WCF overlooks the impossibility for any single form of feedback to be effective across the differently acquired domains of morphology, syntax, and lexis, particularly with respect to grammatical features that are “integral parts of a complex system” (Truscott, 2007, p. 258) that he argued to be impervious to feedback. Truscott admitted the possibility that WCF might prove beneficial for some nongrammatical errors such as spelling or other such discrete errors, but not for errors in grammar (1999, 2007). He also argued that benefits from WCF might be perceived in “pseudo-learning” or at best, self-editing and revision skills without fostering true language development, as it might only develop explicit declarative knowledge (Ellis, 2004), but never implicit procedural knowledge that is argued necessary for acquisition. Truscott’s theoretical arguments were convincing, particularly in translating them to practical concerns. If certain error types could only be amendable to certain types of feedback and/or at certain developmental stages, then the practicality of tailoring WCF to learners’ current level of L2 development must be put to question. The present study, in part, aims to investigate this concern by attempting to gauge a potential learner characteristic (grammatical knowledge) that may influence receptivity to different feedback strategies.

In response to Truscott's claims, Ferris (1999) proposed an ad hoc pedagogical distinction between "treatable" and "untreatable" error types. She defined *treatable* errors as those related to a linguistic structure that occur in a rule-governed way in which the learner can be pointed to a grammar rule (or set of rules) to resolve the problem (e.g., verb tense, subject-verb agreement, singular-plural noun agreement, article usage, run-on sentences, some errors in punctuation or spelling). An *untreatable* error, on the other hand, is idiosyncratic and would require the use of acquired linguistic knowledge to self-correct (e.g., most word choice and preposition errors, missing words, word order). This distinction contradicts Truscott's claim that grammatical errors are impervious to corrective feedback, instead ascribing amenability to feedback based on the extent to which error types are rule-governed, arguing that structures are more amendable to feedback if learners could draw on rules to make corrections. The extent to which this theory of treatability extends from revision to long-term development is unclear, but several scholars have suggested that direct feedback may prove most helpful for errors that do not follow clear rule-based patterns (Chaney, 1999; Ferris, 1999, 2006; Hendrickson, 1980; Park et al., 2016), although only a few studies speak to these claims empirically. Table 1 lists the features that have been investigated in studies that have operationalized the distinction between treatability of error types based on structural characteristics (Ferris, 2006; Park et al., 2016; van Beuningen et al., 2012) and how they were categorized. Ferris et al. (2000) found development over a semester in verb tense and form ("treatable" errors), slight development in lexical- and noun-ending errors such as singular-plural agreement ("treatable"), but regression in the sentence structure (e.g., run-ons, "untreatable") and article categories ("treatable"). In a later study that focused on error type but only in revision, Ferris (2006) found in post hoc analysis that teachers were more likely to intuitively supply indirect feedback for treatable error types and direct

Table 1

*Studies Investigating Truscott's (2007) Error Type "Treatability" Claim*

Study	Dichotomy	Error types included in the study	Treatment suggested (based on findings)
Ferris et al. (2001)	"Treatable" Errors (rule-governed)	verb tense / form, subject-verb agreement, run-on sentences, fragments, noun endings, articles, pronouns	Indirect
	"Untreatable" Errors (more idiosyncratic)	Word choice, sentence structure, lexical errors	Direct
Park et al. (2016)	"Treatable" Errors	Orthography (Korean particles)	Indirect
	"Untreatable" Errors	Lexical	Direct
van Beuningen et al. (2012)	Grammatical Errors	articles, pronominal errors, verb tense, singular-plural noun agreement, word order, fragments, addition or omission of a word, inflectional errors	Direct
	Nongrammatical Errors	Orthographical (spelling, punctuation, capitalization); Lexical (word choice); Pragmatic	Indirect

feedback for untreatable error types, even though the teachers in the study had initially agreed to provide coded indirect feedback to all errors. This observation was notable, revealing the engagement of complex decision-making in the provision of WCF types.

Van Beuningen et al. (2012) investigated Truscott's (2007) categorization of grammatical vs. nongrammatical error types by grouping them and comparing the effects in response to direct

and indirect feedback types. They compared groups of 134 students of Dutch as a foreign language and used a pre-, post-, and delayed posttest design but provided feedback only on a single writing task. They found that different feedback types have different relative value for different types of errors. Only direct feedback promoted durable improvement (medium magnitude) in accuracy for grammatical errors, whereas nongrammatical accuracy benefited most from indirect feedback because it was retained in the delayed posttest. These findings contradict previous accounts that suggest the value of indirect for rule-governed structures (e.g., Ferris, 2006).

Bitchener, Young, and Cameron (2005) also investigated error type, although not operationalized around the notion of treatability. They compared three WCF types (direct, direct with individual conferences, and no feedback) on three error types (prepositions, the past simple tense, and the definite article) over 12 weeks and found development in accuracy for the direct feedback with conference group for past simple tense and the definite article, but no significant effect for feedback in general when the error categories were combined. This finding supports Ferris' (1999) hypothesis of treatability as prepositions are idiosyncratic, and shows the potential for drastic differences in findings depending on the error types under investigation.

A few other studies have focused on linguistic target in students' ability to revise the same writings (e.g., Ferris 2006; Ferris & Roberts, 2001). While development over time in new writings is the ultimate goal of WCF, Ferris (2004) rightfully argues that successful revision is a necessary first step, and indirect feedback on errors for which students are less able to successfully self-edit initially in revision would ultimately be less useful in learning. Ferris and Roberts (2001) focused on revision success rates for ESL students and found them to be more successful at self-editing errors in the treatable categories (verbs, noun endings, and articles) than

the untreatable categories (word choice and sentence structure), although statistical differences lay mostly in the sentence structure category. Likewise, Ferris (2006) found that ESL students were generally successful in editing all error types except for idioms and subject-verb agreements errors.

The classification and grouping of error categories is rarely consistent in WCF research with some studies focusing on structural differences such as lexical, morphological, and syntactic errors (e.g., van Beuningen et al., 2012), while others have considered structural distinctions at the word, clausal, and phrasal levels (e.g., Liu, 2016, that found indirect WCF to be most effective in treating word- and phrase- level errors at the clausal level for editing and also in subsequent development). As findings in this domain have recently provided more definitive support for the general effectiveness of WCF (e.g., Kang & Han, 2015), more attention on the role of linguistic target is needed across classroom contexts and in light of individual learner differences.

**Developmental Readiness.** A critique of the potential value of WCF from an SLA perspective (e.g., Pienemann, 1998; VanPatten, 2007) is that learners could only benefit from WCF on structures for which they are ready to make shifts in development (Truscott, 2007; Ferris & Roberts, 2001). The role of developmental readiness in corrective feedback research has received scant attention. One oral feedback study that has empirically examined the role of developmental readiness in moderating the effects of corrective feedback is Ammar and Spada (2006). They found that more advanced learners with greater mastery of English question formation benefited more from recasts in producing more accurate question formation. They also investigated the interaction between (oral) feedback type and learners' prior knowledge of the target structure, finding that learners with less knowledge about the English possessive

determiners *his* and *her* benefited more from prompts, while prompts and recasts were equally effective for learners with more previous knowledge about the structure. Ammar and Spada (2006) suggests that prior grammatical knowledge of linguistic structure may play a role in moderating the relative effectiveness of different feedback types, however other variables, such as the nature of linguistic target, may be in play. Li (2010) hypothesized that learners with greater proficiency and greater attentional resources might benefit more from corrective feedback in general, and added that it may be misleading to proclaim certain feedback types more or less effective as each feedback type possesses characteristics that may benefit learners at one proficiency level but not another. A few WCF studies have linked developmental readiness to proficiency (e.g., Park et al., 2016) or educational level (e.g., Ferris, 2006; van Beuningen et al., 2012), but none have examined developmental readiness explicitly at the level of individual measurement.

### **Aptitude-Treatment Interaction Research**

Aptitude-Treatment Interaction (ATI) research, a line of research with origins in educational psychology (see Cronbach, 1967) that has recently gained momentum in SLA, aims to match specific types of instruction with learners who will most benefit based on individual differences (Li, 2013). Early SLA researchers such as Wesche (1981) have explored the extent to which L2 learners' language aptitude (e.g., memory, analytic ability) play a role depending on the teaching approach. Robinson (1997) investigated the correlation between language learning aptitude and different learning conditions (incidental, implicit, and explicit), finding that aptitude correlated with the implicit and explicit conditions but not the incidental. In a similar study, Erlam (2005) found that learners with higher language analytic ability and higher working memory capacity benefited more from input-focused instruction that did not require production,

while deductive instruction that allowed opportunities for output minimized the effect of aptitude variation.

Turning attention to corrective feedback research, Li (2010) proposed that aptitude would correlate differently with the effects of explicit and implicit feedback as the two feedback types implicate different cognitive processes. Li (2013) and Yilmaz (2013) both conducted studies that investigated the role of cognitive factors (working memory and language analytic ability) in comparing relative effectiveness of explicit correction vs. recasts in oral CF. Yilmaz (2013) found that both cognitive factors moderated the effect of feedback and that explicit correction worked better than recasts only for learners with high cognitive ability. Li (2013) found that language analytic ability predicted the effects of implicit feedback (recasts), and that working memory mediated explicit feedback (metalinguistic correction). The goals of these studies match closely with the present work, bringing ATI research to the written feedback domain. Although grammatical knowledge is not a static intrinsic trait, it represents an individual difference that has been argued for many years as one of the several factors that likely moderate the utility of WCF without empirical backing.

### **The Potential Role of Grammatical Knowledge on the Efficacy of WCF Types**

The role of explicit knowledge in L2 learning (e.g., Ellis, 2004) has been a central concern in SLA research and a bulk of research on L2 pedagogy seeks to clarify the roles that explicit and implicit knowledge play in different contexts of language learning and teaching (Norris & Ortega, 2000). In view of corrective feedback, modern psycholinguistic theories point to the role of negative evidence for input-processing and restructuring form-meaning relationships (e.g., VanPatten, 2007). In comparing feedback types, theoretical arguments have defended both direct and indirect approaches to WCF. From the perspective of second language

acquisition theory, as well as that of L2 writing specialists, indirect feedback is argued to invite learners in deeper engagement in guided learning and problem solving (Ferris, 2004; Lalande, 1982; Robinson, 2001). Deeper engagement is argued to promote reflection on existing knowledge (or partially internalized knowledge) that is more likely to promote long-term development. Bitchener and Knoch (2008) add that coded WCF requires learners to engage in guided learning and problem solving that promote the type of conscious reflection that is more likely to lead to long-term acquisition. Coded feedback could also enhance learner autonomy by assigning learners part of the responsibility, fostering strategic competence in self-editing skills (Bitchener & Ferris, 2012). It has also been suggested that coded feedback could be less threatening to students compared to more direct approaches (Van Lier, 1988). On the other hand, researchers have also recognized the limited role that indirect feedback could play in facilitating the acquisition of new linguistic knowledge in features for which learners are not yet developmentally ready (Bitchener & Ferris, 2012; Truscott, 2007). From this perspective, arguments in support of direct WCF note that it reduces confusion learners may experience if they fail to understand the meaning of indirect labels or codes and what to do with them, and provides sufficient information for learners to resolve understanding of more complex errors. Direct feedback also offers more explicit and immediate feedback on hypotheses that learners may test as they engage with feedback (Bitchener & Ferris, 2012; Chandler, 2003; Van Beuningen et al., 2008, 2012).

The hypotheses presented above, that direct and indirect WCF serve different roles depending on learner needs related to their existing linguistic knowledge, provide motivation for the present study. Learner readiness, in terms of requisite linguistic knowledge, has been conceptualized in different ways as a potential moderating variable for the relative effectiveness



of feedback types, although it has rarely been operationalized. Ferris and Hedgcock (2004), Ferris and Roberts (2001), and Park et al. (2016), for instance, hypothesized general level of proficiency as an indicator of developmental readiness for indirect feedback, as lower proficiency learners may not possess the required linguistic knowledge to make corrections that are marked by a code. Several scholars have argued that learners may need a certain level of (meta)linguistic competence to be able to self-correct using indirect WCF (Ferris & Roberts, 2001; Ferris, 2004; Hyland & Hyland, 2006; Li, 2010; Park et al., 2016; Sheen, 2007). Park et al. (2016) cautioned that “while indirect feedback can be a useful and practical technique, its provision should be based on teachers’ careful evaluation of their students’ prior L2-exposure / learning experience and the level of metalinguistic knowledge that different students bring to the classroom” (pp. 695-696). Van Beuningen et al. (2012) discussed learners’ level of metalinguistic awareness in terms of educational level, noting that lower metalinguistic awareness at the secondary school level in their study may have prevented pupils from fully benefiting from indirect WCF they received, although metalinguistic knowledge was not measured directly.

The only published WCF study that explicitly investigates the role of learners’ formal knowledge of grammar on their use of feedback is Ferris and Roberts (2001). However, they focused only on students’ ability to revise in the same writings and did not compare feedback type in relation to grammar knowledge. To connect learners’ prior knowledge of six targeted error types with patterns of student success in revision, they developed a “Grammar Knowledge Pretest” and a “Grammar Knowledge Questionnaire” although the reliability of these measures was not reported. Adult university ESL students were grouped into three feedback groups: indirect (underlined and coded), indirect (underline only), and a control. They correlated pretest

scores on the Grammar Knowledge test with students' error rates in first drafts and with their successful revision in second drafts. No correlation was found with individual error rates in first draft performance, but in terms of revision, success in editing on four out of the six error types (verbs, noun endings, word choice, and total errors) significantly and positively correlated with the Grammar Knowledge pretest scores. Only revision in articles and sentence structure did not correlate with Grammar Knowledge scores. These results suggest that formal knowledge of grammar assists in the revision process in response to indirect feedback. The present study shifts attention to the role of grammatical knowledge on learners' relative utilization of direct and indirect feedback in developing accuracy over time in new writings.

The methods in answering this question follow the Aptitude-Treatment Interaction studies described above (e.g., Li, 2013; Yilmaz, 2013) that attempt to match specific types of instruction with learners who will most benefit based on individual difference characteristics (e.g., learning abilities, learning styles). This line of research is promising to inform our understanding of the interaction between learner-external and learner-internal variables. The present study aims to tease apart the influence of the interaction between prior grammatical knowledge with feedback type on gains in accuracy. The next section of the literature review defines the construct of grammatical knowledge for the purposes of this study and motivates the instrument designed to measure it.

### **Defining and Measuring Grammatical Knowledge**

To build construct validity it is necessary to first identify a conceptualization of grammatical knowledge that can be used as a basis for constructing an assessment instrument and then provide a validity argument for its operationalization. This section of the literature review explores the types of linguistic features investigated in WCF studies, the

conceptualizations of grammatical knowledge in L2 research, and reviews instrumentation previously used in its measure.

**Targeted error types.** Studies investigating the effectiveness of comprehensive WCF focus on a range of error types, categorization for which has been inconsistent between studies and is sometimes left unreported (Liu & Brown, 2015). Those studies that have investigated development in new writing for a wide range of error types (i.e., comprehensive in scope) and that have specified individual error types are listed in Table 2. Error categories that were present in more than three different studies are included in the table. Note that among these 18 error

Table 2

*Error Types Investigated in Previous WCF Studies that Include Comprehensive Feedback*

Studies	VT	WO*	WC*	MW*	SP	SV	*Prep	pos	Art	Frag	Sent Struct*	Sp*	RO	Pro	Punc
Lalande (1982)	x	x	x	x	x	x					x			x	
Chandler (2003)	x	x	x	x	x	x		x	x	x	x	x	x	x	x
Bitchener et al. (2005)	x						x		x						
Rahimi (2009)	x	x	x	x	x	x	x		x	x	x		x	x	
Hartshorn et al. (2010)	x	x	x	x	x	x	x	x	x	x		x	x		x
Benevento & Storch (2011)	x	x	x	x			x	x	x		x	x			
van Beuningen et al. (2012)	x	x	x	x	x		x	x	x	x		x			x
Riazantseva (2012)	x	x	x	x	x		x	x		x	x		x	x	
Sampson (2012)	x	x	x	x	x	x		x				x			x
Totals	9	8	8	8	7	6	6	6	6	5	5	5	4	4	4

(VT = verb tense; WO = word order; WC = word choice; MW = missing word; SP = singular-plural; SV = subject-verb agreement; Prep = preposition; pos = part of speech; Art = articles; Frag = Fragment; Sp = spelling; RO = run-on; Pro = pronouns; Punc = punctuation)

\* Error types that are less rule-governed (i.e., more idiomatic in formation), include word order, word choice, missing word, sentence structure, spelling, and extra word.

types, six (word order, word choice, missing word, sentence structure, spelling, and extra words) have been considered idiomatic and rarely rule-governed (e.g., Ferris, 2006). This dichotomy will be discussed further in the methods section to justify error selection for this study.

In some studies with interest in error type, the selection of target linguistic features has been influenced by the notion of “treatability” or grammaticality of error types (Ferris, 2006; van Beuningen et al., 2012). A valid criticism of WCF raised by Truscott (1996) was that different types of linguistic forms (i.e., morphological, lexical, syntactic) may be variably amenable to corrective feedback because they represent different domains of linguistic knowledge that develop following differing sequences of acquisition. The present study, therefore, includes examples of linguistic features that include morphological, lexical, and syntactic error categories to compare effects.

**Conceptualizing grammatical knowledge.** From a theoretical SLA perspective, the measurement of grammatical knowledge often involves distinguishing between explicit or declarative knowledge and implicit or procedural knowledge (Ellis, 2004; Rebuschat, 2013). Explicit knowledge is defined as conscious understanding of language, generally accessible through controlled processes, potentially verbalizable, and naturally called upon for exploitation when a learner faces difficulty in a language task. Implicit language knowledge, on the other hand, is entirely tacit but available for automatic use (Ellis, 2004; Rebuschat, 2013). Explicit knowledge has been studied from various perspectives using terms such as *metalinguistic awareness / ability / performance*, or *analyzed / conscious/ learned / declarative / explicit knowledge*. As Ellis (2004) points out, grammatical knowledge is often associated with explicit knowledge in SLA research, perhaps because in contrast to pronunciation and vocabulary, it is more amenable to conscious reflection and manipulation and there is also a well-established

metalanguage for the discussion of grammar. Grammatical knowledge in the present study encompasses explicit knowledge in addition to elements of implicit knowledge introduced below.

**Grammatical knowledge in the context of WCF.** The measurement of grammatical knowledge for the purpose of understanding how learners utilize WCF has received limited attention (see Ferris & Roberts, 2001, for an exception). In this study, grammatical knowledge is conceptualized as the linguistic knowledge that can potentially influence the utility of indirect form-focused WCF. This knowledge can be either implicit or explicit, and encompasses the ability to recognize the metalanguage that identifies an error type (i.e., the code) and to understand the form-function relationship of targeted linguistic structures and the rules associated with their use. The following section explores the instrumentation that has been employed in previous research to measure L2 grammatical knowledge to motivate the methods in the design of the measure developed for this study.

**Methods of assessing grammatical knowledge.** Due to the lack of consensus in L2 research about what constitutes grammatical knowledge, necessarily much disagreement follows in terms of how to assess it (Purpura, 2004). The traditional approaches to assessing grammar involve the measurement of accurate production and comprehension by means of discrete, decontextualized items in formats such as sentence completion, error correction, and judgment of grammatical correctness (Larsen-Freeman, 2009; Purpura, 2004). More recent approaches to the assessment of grammatical knowledge focus on the ability to integrate and use knowledge of grammar in performance, often assessed holistically and subjectively through speaking and writing tasks by raters using scales that gauge grammatical accuracy and complexity (Purpura, 2006). This dichotomy between discrete-point and integrative assessment represent different

construct definitions, approaches, and purposes. While it is recognized that selected response tasks that test knowledge of forms in isolation may fail to capture the dynamic and complex understanding of the resources needed for communication, they can nonetheless be useful in assessing knowledge of isolated features, particularly for research purposes (Purpura, 2013).

Explicit knowledge in L2 research is most commonly measured by grammatical judgment tasks (GJTs), error correction tasks, and tests of metalinguistic knowledge or awareness, among others. Table 3, adapted, expanded and updated from Ellis (2004), summarizes the tasks included in several studies that have measured explicit grammatical knowledge in L2 research. In judging the suitability of GJTs, their reliability has been called into question (Birdsong, 1989) due to the dangers of response bias (e.g., the general tendency to judge sentences as ungrammatical), which is often addressed with the use of distractors. It should also be noted that error correction tasks can be considered to tap into explicit or implicit knowledge. Time has shown to be a crucial factor in distinguishing between implicit and explicit knowledge in GJTs in Godroid, Loewen, Jung, Park, Gass, and Ellis (2015), as learners can be expected to rely more on implicit knowledge under time constraints but the opportunity of more time can open access to explicit knowledge.

**Defining and operationalizing metalinguistic knowledge.** The construct of metalinguistic knowledge (MK) has been defined in different ways in L2 literature, but overlapping themes point to the notion of explicit knowledge about language and the ability to express such knowledge. MK is of interest in this study, specifically, because it has been hypothesized as a moderating variable of WCF (e.g., Bitchener & Ferris, 2012; Hyland & Hyland, 2006; Park et al., 2016). MK is rule-based, declarative (rather than procedural),

consciously analyzed, and requires deliberate focus (Roehr & Gánem-Gutiérrez, 2009). It

involves the ability

to manipulate and explain the attributes of language. Learners are aware when they are drawing

Table 3

*Instruments Designed to Measure Explicit L2 Grammatical Knowledge* (expanded from Ellis, 2004, p. 246-248)

Study	GJT	Error ID	Error Correction	Rule description / explanation	Other
Bialystok (1979)	x	x			Choose (from a list) which rule has been violated
Sorace (1985)	x	x	x	x	
Masny (1987)	x		x	x	
Green & Hecht (1992)	x		x	x	Grammaticality explanation test
Alderson et al (1997)	x		x	x	Metalinguistic assessment test for English and French; Identification of named parts of speech
Han and Ellis (1998)	x			x	
Elder, et al (1999)	x		x	x	Metalinguistic assessment test for English and French; Identification of named parts of speech
Butler (2002)				x	Verbalization of reasoning for choosing article forms in a cloze task
Clapham (2001)					Identification of metalinguistic terms and ability to use them varied according to whether the sentence was simple/complex
Hu (2002)	x		x	x	Judgment of the prototypicality of the rules

on MK (in judging the grammaticality of a sentence, for example), whereas implicit or automated knowledge may be drawn on unconsciously in judgment about language or in the context of language use (Elder, 2009; Ellis, 2004; Roehr & Gánem-Gutiérrez, 2013). A common related term is metalinguistic *awareness*, which has been defined as “an individual’s ability to match, intuitively, spoken or written utterances with his or her knowledge of language” (Masny, 1987, p. 59). Metalinguistic awareness is distinct from MK in that it can involve implicit rather than (or in addition to) explicit knowledge, while MK can be considered analytical in that it involves explicit declarative facts that are known about language (Elder, 2009).

Regarding its operationalization and measurement, MK has been examined in a variety of ways in L2 research. It is usually examined using a combination of measures such as identification of grammatical constituents of sentences (Correa, 2011), GJTs (e.g., Akakura, 2012), cloze test judgments (Serrano, 2011), identification and/or correction of errors (e.g., Akakura, 2012), ability to make comparisons between different L2s (Tellier & Roehr-Brackin, 2013), identification of errors (Sorace, 1985), and verbalization of rules (e.g., Correa, 2011; Gutiérrez, 2013; Han & Ellis, 1998). Table 4 outlines the studies that have incorporated these common formats. MK is usually measured using a combination of measures, frequently including grammaticality judgment, error correction, and verbal explanation (e.g., Alderson, Clapham, Steel, 1997; Hu, 2002; Roehr & Gánem-Gutiérrez, 2009). The most common operationalization of MK in L2 studies (shown in Table 4) involves the explicit verbalization of rules, usually in writing but also elicited orally. Several kinds of verbal reports have been used, such as provision of explanation of grammatical judgments (e.g., Butler, 2002), or description of why a sentence is grammatical or ungrammatical (Green & Hecht, 1992). Hu (2002) acknowledges, however, that verbal report tasks may not be an exhaustive measure of knowledge



Table 4

*Metalinguistic Knowledge Test Designs*

<b>Study</b>	<b>Test purpose</b>	<b>Test Format</b>	<b>Target Features</b>	<b>Scoring</b>	<b>Reliability</b>
Sorace (1985)	Determine the relationship between MK and language use.	Identify ungrammatical sentences, correct them, and state the grammatical rules that had been broken, no time constraint.	Italian perfect, imperfect, indirect pronoun, auxiliary choice, past participle agreement	(0-3 points) 1 for identifying the error, 1 for making a correction, 1 for stating the correct rule.	Not reported
Alderson et al. (1997)	Determine the relationship between MK and proficiency / aptitude	Identify ungrammatical sentences, make corrections, and state the grammatical rule that had been broken	English sing-plural, passive verb, article, prep., rel. pron., aux verb, past participle, conjunction, finite/infinite	(0-3 points) 1 for accurate correction, 1 for producing an accurate rule, 1 for giving the reason.	Items were piloted and “only ones which were statistically satisfactory were retained”
Elder, Erlam, & Philp, 2007)	To determine the MK of advanced EFL learners and teachers in Malaysia	15 ungrammatical sentences, each with a typical learner error underlined for written explanatory rule	English modals, articles, adverb placement, poss, quest. formation, past tense, S-P, conditional, comparatives	(0-1 point) 1 point for an adequate explanation.	Cronbach’s alpha = .81
Hu (2002)		12 Written explanation in Chinese.	English articles, subject-verb agreement, verb tense, aspect	(0-1) correct or incorrect	Interrater agreement of 96%
Akakura (2012)	To measure the effectiveness of explicit instruction on implicit/explicit knowledge.	Correct 10 sentences and provide written explanation for the ungrammatical elements ( $n = 5$ ); no time constraint, two practice items	English articles	(0-1) correct or incorrect	Croncach’s alpha = .64
Gutiérrez (2013)	To determine the relationship between MK and metalingual knowledge and their relationship with proficiency	16 sentences with an underlined error, provide written description of the rule being violated.	Spanish: det. agreement, SV, stem-changing Vs, irregular Vs, ser & estar, imperfect, preterite, subjunctive	General rubric (0-6 points) to score all structures	Cronbach’s alpha = .85
Spada, Jessop, Tomita, Suzuki & Valeo, 2014	To determine the contribution of form-focused instruction on implicit/ explicit knowledge	24 items, first error identification, then error correction, then error explanation.	English passive construction (use of present participle, absence of aux, use of bare form).	1 pt. identify, 2 pts. correction, 2 points explanation	Reported in IRIS as “high levels of reliability obtained”

because learners may have the explicit knowledge of a specific rule but also lack the ability to verbalize the rule. This limitation suggests that verbal reports should ideally be accompanied by more receptive tests as well. Several have also acknowledged the inherent difficulty in measuring the ability to provide explicit grammatical rules because learners' verbalizations vary on a continuum of accuracy and precision. This continuum has been reflected in the development of rubrics that often include several points (up to six) for rating verbal reports (e.g., Gutiérrez, 2013; Correa, 2011; Han & Ellis, 1998; Serrano, 2011). In summary, the measurement of MK has involved a mixed approach in the measurement of explicit knowledge of linguistic form, and often extends to tap into knowledge of technical grammatical language. The following section describes the scoring procedures that have been employed in measuring the most common test format, rule explanation.

**Scoring criteria for metalinguistic knowledge.** Most studies that have measured MK with a rule explanation task incorporate a generalized scoring rubric to rate responses on a scale of accuracy, judging both the substantive content of rules and test-takers' use of metalingual terminology. The level of specificity of these rubrics ranges between scales with three levels (Correa, 2011), four levels (Serrano, 2011), five levels (Wistener, 2014), and six levels (Gutiérrez, 2013; Han & Ellis, 1998). Note in the example rubric presented in Table 5 from Gutiérrez (2013) that the scale is designed to score an explanation for any grammatical error type and includes reference to technical metalanguage.

Although several rubrics that measure rule explanation have relied on a single, generic scoring system to be applied across target error categories, shortcomings are apparent in this method, particularly when a wider range of error categories are investigated. A challenge here lies in the inherent disparity that exists between error types, as the significance of MK differs

Table 5

*Example Scoring Scale for the MK Rule Explanation (from Gutiérrez, 2013, p. 180)*

Score	Description
0	The learner does not verbalize the rule nor corrects the error
1	The learner corrects the error, but does not verbalize the rule
2	The learner verbalizes the rule using some technical metalanguage, but the rule is incorrect.
3	The learner verbalizes the rule using technical metalanguage, but the rule is imprecise, incomplete, or partially correct.
4	The learner verbalizes the correct rule, but does not use any technical metalanguage
5	The learner verbalizes the correct rule using some technical metalanguage
6	The learner verbalizes the correct rule using appropriate technical metalanguage

between linguistic features depending on the variability within each category and how clearly the description of the construction of a grammatical feature matches a pedagogically formulated rule (Robinson, 1997; Roehr & Gánem-Gutiérrez, 2009; Zietek & Roehr, 2011). A few of the studies that have measured MK with rule explanations prove exceptions (Elder et al., 2007; Spada, Jessop, Tomita, Suzuki, & Valeo, 2014) by providing an answer key that specifies the operationalization of MK for each distinct error type. Table 6 presents two examples from Elder et al. (2007) that illustrate the distinct criteria within each error type:

Table 6

*Examples of Scoring Criteria for Specific Error Categories (from Elder et al., 2007)*

Example #1- Singular-plural noun agreement

---

Example Error	<i>She grew <u>some rose</u> in her garden.</i>
1 point	the need to pluralize particular types of noun after ‘some’
1 point	the use of the words: countable / plural / quantifier
2 points (example of a totally correct explanation)	The noun is countable, so after “some” use the plural form.

---

Example #2- Articles

---

Example error	Because he was late, he called <u>taxi</u> .
1 point	reference to the idea that you need a determiner/specifier before a noun.
1 point	The use of the words: (indefinite) article/determiner/countable/singular noun/specifier
2 points (example of a totally correct rule)	You need an article/determiner before a countable noun. Use ‘a’ here because it is a non-specific/indefinite taxi.

---

**Disparity between error types.** The notion that characteristics of MK differ between error categories has been theorized in SLA research in view of prototype theory, which accounts for how categories are cognitively represented and processed (Rosch & Mervis, 1975). Prototype theory explains that any category consists of a prototype structure with some members being more prototypical (i.e., better examples) of the category than others, suggesting asymmetries among category members and asymmetric structures within categories. Hu (2002) reviewed the relevance of prototype theory to research on MK in L2 learning, concluding that:

The notion of linguistic categories as prototype structures has important implications for research on the role of metalinguistic knowledge in L2 production. Because metalinguistic knowledge is explicit knowledge of linguistic categories and form-

meaning relations within categories, it is reasonable to hypothesize that acquisition and use of such knowledge can be influenced by the inner structure of these categories. Additionally, pervasive prototypicality effects, such as ease of learning, order of acquisition, and frequency of item output, may have further impacts on both the content of metalinguistic knowledge acquired and the development of executive control over such knowledge. (p. 335)

Hu's (2002) study discovered that MK was constrained several psychological factors and found significantly greater grammatical accuracy for more prototypical uses in output on writing and error correction tasks. Linguistic prototypicality has also been addressed in L2 research to operationalize explicit learning difficulty for different linguistic structures. Roehr and Gánem-Gutiérrez (2009) developed a taxonomy to judge explicit learning difficulty, rating grammatical constructions as low in explicit learning difficulty (such as third person –s in subject-verb agreement) when they are high in schematicity (i.e., the extent of prototypical use relative to peripheral), low in conceptual complexity (i.e., the rule is a simple proposition consisting of two categories), low in technicality of necessary metalanguage (e.g., “subject,” “third-person”), and high in truth value (i.e., the rule applies without exception). Compared to subject-verb agreement, which is highly prototypical in use, verb tense can represent a category with a range of prototypical uses, such as distinctions between past, conditional, and future tenses, with peripheral uses such as irregular formations. Likewise, the difficulties that L2 learners face in learning the complex rule structure associated with English articles is well documented, and the idiomatic nature of prepositions limits the ability of learners to develop conscious form-meaning mappings for such features.

From the perspective of explicit learning difficulty, the relevance of MK for each error category must be viewed as distinct in that different error types vary in their internal variability of form-meaning relationship, the extent that idiomatic item-learning is required rather than rule-based learning, and in the range of peripheral uses beyond prototypical use. Therefore, in view of measuring MK, the more we can account for the qualitative distinctions within each error type, the better. And these distinctions can be considered both in terms of internal structural and in terms of representativeness in real language use.

### **Summary of the Literature Review**

In summary, the present study aims to provide further evidence of the effectiveness (or lack thereof) of WCF in an EFL context. The relative effectiveness of direct and indirect WCF types will be compared overall and at the level of error type. The unique contribution of the present study is the exploration of the potential role of grammatical knowledge on L2 learners' use of WCF. This variable often has been hypothesized to moderate learners' use of different feedback types, but has not been studied empirically with the dependent variable as the development of accuracy over time. The measure of grammatical knowledge developed for this study is tailored to the linguistic performance of the L2 writers in the target context and draws on common approaches outlined in previous literature that include both explicit and implicit measures of grammatical knowledge argued here to play a role in helping learners engage with and learn from the feedback they receive. Grammatical knowledge, for the purposes of this study, is operationalized through the sub-constructs of error type identification, guided error correction, and metalinguistic rule description. With clearer understanding of the relationship between grammatical knowledge and the relative efficacy of direct and indirect feedback types, practitioners could be better informed in decisions of which types of feedback to provide to

exploit the existent knowledge that students bring to a writing classroom. The following four research questions were formed to address these concerns:

### **Research Questions**

**RQ 1.** To what extent are direct and indirect WCF effective in developing learners' accuracy, relative to each other and to no WCF?

**RQ 2.** To what extent are the different error types relatively amenable to WCF in developing learners' accuracy in response to direct and indirect feedback types?

**RQ 3.** To what extent does grammatical knowledge moderate the effectiveness of WCF in developing grammatical accuracy?

**RQ 4.** To what extent does grammatical knowledge of each error type interact with feedback type on learners' development in grammatical accuracy?

## CHAPTER 3

### METHODOLOGY

This chapter presents the methods including the general research design, variables in the study, research setting, participants, writing assignments, selection and rationale for the target error categories, details of the treatments, development of the Grammatical Knowledge Test, data collection procedures, and the methods of analysis.

#### **General Research Design**

The quasi-experimental research design included three pre-existing groups randomly assigned for direct ( $n = 41$ ) and indirect ( $n = 44$ ) feedback, and a third group as a contrast group ( $n = 42$ ) that received feedback on content and organization but not on accuracy. Data was collected from classes in two consecutive academic quarters in a Thai EFL context and then combined into a single data set. The study uses a pretest-posttest design to determine the gains in written grammatical accuracy.

#### **Variables in the Study**

There are three main variables in the study: one dependent variable and two independent variables, illustrated in Table 7. The dependent variable is the gain in accuracy (overall and within each error type), measured by subtracting the normed frequency of errors from the posttest to the pretest. The first independent variable (often termed predictor variable, PV, or explanatory variable in multivariate analysis) is the grammatical knowledge of the structures that represent each error type, measured by a test administered at the start of the course. The data for the Grammatical Knowledge Test scores is continuous and represents the composite scores of three sub-sections of the Grammatical Knowledge Test. The Grammatical Knowledge Test scores and the gain scores for improved written accuracy were converted to *Z*-scores to serve as



Table 7

*Description of Study Variables*

Category	Type	Variable
DV	Continuous	Gains in accuracy (development within each error type individually) between the first drafts of the first and final in-class essays
IV #1	Continuous	Composite Grammatical Knowledge Test score for each grammatical feature
IV #2	Categorical (3 levels)	Feedback/Treatment type (1. indirect, 2. direct, and 3. control)

a standard measure of dispersion so that they could be comparable to investigate Research Questions 3 and 4, even though their scales are different. The second independent variable is WCF type, (direct or indirect, in comparison to a contrast group). While WCF type and gains in accuracy have received significant attention, the novel aim of this study is to shed light on whether grammatical knowledge predicts the value of one treatment type relative to the other on improving gains in accuracy. In other words, if learners score high/low for grammatical knowledge for certain linguistic features, would indirect or direct WCF prove more beneficial in treating those features?

**Research Setting**

The study took place at an intensive English program (IEP) at an elite university in Thailand. The IEP allows promising applicants to the university who are not fully prepared in

academic English to be accepted contingent on graduating the IEP program within their first year. The program therefore bolsters students' academic English skills to prepare them for full-time coursework in English in the international school at the university. Students in the IEP are highly motivated to improve their academic language skills to matriculate into the university. Students' readiness to matriculate into the university was assessed in terms of students' listening, speaking, reading, and writing abilities, however, their writing scores were most heavily weighted in the overall assessment that determines progression through the four levels of the program, and completion of the program after level 4. In addition, accuracy played a significant factor in the scoring rubric on their high-stakes assessment.

Designing a WCF study with a control group can present an ethical challenge if stakeholders assume that withholding WCF neglects a control group. This has been a cause of methodological inconsistency in this line of research and proved a dilemma in the context of the IEP in this study. To set the parameters of the study so that a control group could be included that did not receive WCF on grammar, and so that students in any of the three groups would not receive explicit grammar instruction on the target features throughout the course of the study, it was agreed that the study would only last seven weeks out of the 10 week quarters. This allowed the final three weeks (following the posttest) for teachers to focus grammar instruction on the features targeted in this study and for the control group to receive WCF on grammar for all their writings that was withheld throughout the study. The constrictions in this design to match curricular goals led to the decision not to include a delayed posttest in favor of maximizing treatment time.

All students in the IEP were required to bring laptops to every class and all classes used the Google Classroom platform, allowing for assignments to be consistently written on Google

Docs. The researcher, also the teacher of the control groups, had access to the GoogleDrive writing folders of all students in the study so that WCF could be provided digitally in a timely manner. Classes in the IEP are divided by skill and set at four levels of proficiency. Most students place into the program at level two or three, and then must progress through level four to matriculate into the international school. They are permitted to fail a level only once (in consecutive quarters) to remain in the IEP, and about 15-20% of students fail each level and are required to repeat. IRB approval (Appendix E) was obtained for the study.

### **Participants and Grouping**

The 127 participants in this study were 61% female, 39% male with an average age of 18.8. The majority of students ( $n = 115$ ) were of Thai nationality and had lived in Thailand, although about 10% of these students had experience studying abroad in a predominantly English speaking country for varying lengths of time (from a few weeks to a year). The remaining 12 participants were international students, mostly from the surrounding region (three Japanese, two Taiwanese, two Indian, two Korean, and one Chinese), and two Thai students had completed high school in the U.S. with spoken fluency in English but limited ability in academic English writing. Students aim to reach low-advanced proficiency upon completion of level 4, which allows them to matriculate into the university. Students are randomly assigned to one of the 4-6 sections within each level. The study was conducted in level 3 of the IEP, as the upper level curricula included substantially more writing assignments of greater length than the lower levels, which allowed for a sufficient quantity of writing within the seven weeks of data collection. Level 3 was selected rather than 4 because it was considered slightly lower stakes with respect to the ethical concerns of including a control group, since completion of level 4

determines students' acceptance into the university's international school and accuracy in writing could potentially prevent a student from passing through the program.

The IEP placement test was the same test used for the writing achievement test and determined which level students begin and successfully complete. To progress through levels, students must pass each skill course, although their scores on a timed essay (similar in style to IELTS writing section with unreferenced in-class essays) weigh most heavily in the overall assessment. Although an in-house rubric was used to determine placement and advancement through levels, for the purposes of this study two raters with previous training as IELTS raters (one had experience as an IELTS trainer) were enlisted to re-score the writings used for placement into level 3 (or for advancement from level 2 to 3) using the public version of the IELTS rubric that uses a scale of 1-9. These writing tests were administered two weeks before students began level 3. The scoring was done on approximately half of the data set (i.e., the first term in which the study was conducted,  $n = 67$ ). If the two raters disagreed by more than 1 point, a third rater's score was used. Ratings were averaged together for a final writing proficiency pretest score. Results showed that at least 84% of students in each group scored between IELTS 5.0-6.5 (direct:  $M = 5.37$ ,  $SD = .48$ , indirect:  $M = 5.52$ ,  $SD = .57$ , control:  $M = 5.47$ ,  $SD = .63$ ), scores that are equivalent to bands B1-B2 on the Common European Framework, or 16-23 on the TOEFL-iBT writing section. A one-way ANOVA conducted on the scores with group condition as a between-subjects variable (significance set to .05) showed that there were no significant differences among the three groups in terms of proficiency in writing at the onset of the study,  $F(2, 66) = .242$ ,  $p = .786$

## Writing Tasks

Students in level 3 write, on average, one essay every week and these writing tasks were designed to prepare students for the mid-term and final exams that were similar in style to the IELTS writing section. The writings in level 3 typically included both referenced and unreferenced academic essays, although for the purposes of the study, the referenced essay practice was reserved for the final weeks of the course (following completion of the data collection) so that the study could control for writing type. Unreferenced writings were chosen as they would limit paraphrasing and citation as potentially confounding variables. Writing topics in the level 3 writing course were generally argumentative academic topics related to local or social issues in Thailand or the university. Topics for the writings used in this study were consistent across sections and agreed on by the teachers of the three classes involved in the study. These topics were chosen lists of practice IELTS writing prompts and had been used in previous terms so that instructors felt comfortable regarding their level-appropriateness. The topics were then repeated for the subsequent quarter to double the data set. The topics are listed in Table 8.

Table 8

### *Essay Topics*

Role in study	#	Topic
Pretest (first half)	1	Should children be allowed to use mobile phones?
Pretest (second half)	2	Should university students have part-time jobs?
<i>and</i> Treatment 1		
Treatment 2	3	Should women be required to do military service in Thailand as well as men?
Treatment 3	4	Should animals be kept in zoos?

Treatment 4	5	Should gambling be legal in Thailand?
Treatment 5	6	Should there be a tax on unhealthy food in Thailand?
Posttest (first half)	7	Should the use of pesticides and/or GMOs be banned in Thailand?
Posttest (second half)	8	Should high school students be required to bring laptops to class?

All eight of these in-class essays served as practice essays in preparation for the mid-term and final exam writing tests. The mid-term essay (administered in week five) did not receive feedback and was not included in this analysis. All writing tasks were timed and completed within the 1 hour and 40 minute class period under the same conditions. Writings were done in-class because of anecdotal evidence of much more variation in the students' take-home work than in their in-class work, as some students enlist the help of parents/ siblings/ friends with their take-home work and some devote much more time than others outside of class. For each writing, students were advised to sketch an outline for the first 5-10 minutes of the class period before typing an essay of at least 350 words. They were also instructed to avoid use of any resources on the web to mirror authentic testing conditions students would face on the exams. They were not required to finish the essay, but the vast majority were able to exceed the minimum 350 word requirement.

After students made revision in response to corrective feedback, they received feedback on content and organization on a second draft that was submitted and graded for practice. The rubrics to grade the practice essays incorporated a band for holistic accuracy weighing one-third of the grade. Although the scores students received on essays in the pretest, treatments, and posttest did not directly affect their grade in the course, students were required to complete all the essays and these practice essays were taken seriously as students realized they needed to

practice to pass both the midterm and the final exams, which included a timed unreferenced essay similar in style to the in-class practice.

### **Selection and Rationale for the Target Error Categories**

Rationale for the selection of the target linguistic features used in this study are described below, generally adhering to the principles of representation, ecological validity, and susceptibility to measurement of metalinguistic knowledge. The features selected for this study are described in Table 9 and details of the selection criteria are described below:

**Comprehensive without overwhelming in scope.** The aim was to include a wide scope

Table 9

#### *Description of Error Categories and Codes Used for Feedback and Analysis*

Error type	Code*	Description / Notes
Singular-Plural Agreement	SP	A singular noun is used where a plural noun is required or vice versa. Excludes misspelling or errors in irregular plural forms.
Articles	Art	An article is omitted where it is required; an article is used when none is called for, a/an/the are incorrectly used.
Subject-Verb Agreement	SV	A verb does not agree in number or person with its subject.
Part of Speech / Word form	POS	A word is used as an incorrect part of speech or word family (i.e., inappropriate noun, verb, adjective, or adverb).
Verb Tense/Aspect	VT	Inappropriate tense or aspect is used for the context. Tenses include errors in past, present, and future modal use. Aspect includes errors in progressive and perfect. Complex structures such as voice (active vs passive) and conditional are excluded.
Run-On Sentence	RO	Independent clauses are joined without appropriate punctuation or conjunction.

\*For indirect feedback group only.

of error types that cover the majority of errors made by students in the target context in order to preserve ecological validity. However, another goal was to avoid overwhelming students with too many different error types, which may cause difficulty for students in understanding or recognizing distinct error categories. The six error types selected comprise 71.9% of the total errors found in an error analysis of a sample of 40 level 3 student essays (605 out of 841 total errors) that were written in the previous year.

**Rule-governed structures.** To measure metalinguistic knowledge as a variable in this study the features targeted are (somewhat) rule-governed and less idiomatic in their form-meaning relationship. Error categories need to be descriptively accessible in that it should be possible to elicit metalinguistic knowledge in the process of constructing appropriate forms or making corrections. Ferris (2006) termed these error types as “treatable,” while Van Bueningen et al. (2012) referred to such error types as “grammatical.” “Non-grammatical” or “untreatable” error types, such as word choice or spelling errors were not targeted (Table 10 below labels examples of idiomatic error types with asterisks).

**Representative of different categories of structures.** Selected error categories represent features from morphological (verb tense/aspect, subject-verb agreement, and singular-plural agreement), lexical (articles, parts of speech/ word family), and syntactic errors (run-on sentences) to include representation from each of the three overarching categories of language use.

**Most prominent for the target population.** An error analysis was conducted on a sample of previous in-class writings of 40 essays (20,392 words), each written by different students in the previous year from two sections in level 3. These essays were individual, timed, and unreferenced on academic topics from four different prompts (e.g., *Is it important for*



*parents to be strict with their children or allow them more freedom?*), mirroring the target writing context for data collection in this study, although the exact topics were not duplicated. This analysis informs the selection of targeted categories with the aim of including most level 3 students' errors for treatment and analysis. Table 10 presents the 13 most common error types

Table 10

*Most Frequent Error Types in Level 3 Writing (40 essays, 4 different prompts, 20,392 words)*

Rank	Error Type	Raw Frequency	per 1,000 words
1	<b>Singular-Plural Noun Agreement</b>	193	9.46
2	<b>Articles</b>	167	8.19
3	<b>Subject-Verb Agreement</b>	84	4.12
4	<b>Part of speech</b>	66	3.24
5	*Prepositions	64	3.14
6	<b>Verb Tense / Aspect</b>	58	2.84
7	*Word Choice	43	2.10
8	<b>Run-on sentences</b>	43	2.10
9	Complement clause (gerund / infinitive)	27	1.32
10	Pronouns	26	1.28
11	Verb voice (active / passive)	17	.83
12	*Verb omission	14	.69
13	Sentence Fragment	11	.54

\* Errors that are less rule-governed and more idiomatic in their form-meaning relationship.

found in the sample. Note that the error types deemed to be more idiomatic in nature, and therefore less amenable to metalinguistic knowledge, are labeled with asterisks. Verb voice (#11

most frequent) was considered too challenging for these upper-intermediate learners and therefore separated from the verb tense/aspect category. The six targeted error types selected for this study (highlighted in bold in Table 10) were all among of the eight most frequent found in level 3 student writing and occurred at least once per essay (i.e., based on an average essay length of about 500 words).

**Representative of previous research.** Previous studies of WCF were referenced to ensure that the error types selected in this study would retain generalizability and comparability to the broader domain. As discussed in the literature review, Table 2 presents a summary of error types in all published studies (through 2014) that have investigated development over time, that have included comprehensive (or somewhat comprehensive), indirect feedback, and that have reported individual error types. This table overlaps considerably with both the error analysis conducted on the level 3 essays in the target context and with a previous analysis of Thai EFL university writing. Pongsiriwet's (2001) dissertation was used to triangulate the analysis conducted on level 3 student writing to ensure ecological validity for the target population. Pongsiriwet's (2001) dissertation investigated compositions written in-class by 155 university freshman in an EFL context in Thailand. These essays described or narrated a story about significant events or people in students' lives. Pongsiriwet's error analysis of this larger sample of writings identified the ten most frequent types of grammatical errors (percentages of overall errors in parentheses) as the following (error types in bold are included in the current study):

1. **subject-verb agreement** (14%);
2. verb formation (13.7%) – this category includes spelling errors within verb phrases and errors in modal usage;
3. **verb tense** (13.1%)

4. **singular-plural noun agreement** (11.7%)
5. **part of speech / word form** (11.1%)
6. verb omission (9.9%)
7. **articles** (7.1%)
8. prepositions (6.7%)
9. sentence fragments (5.9%)
10. pronouns (3.4%)

Pongsiriwet's findings generally overlap with the results of the smaller scale error analysis of level 3 student writing, as five of the six error types selected for this study are reflected in Pongsiriwet's top seven, except run-on sentences, which occurred more frequently in the level 3 writing (perhaps due to more academic/complex prompts) and remain in the present study to provide representation of structural errors.

### **Treatment**

Although teachers in the IEP typically provided WCF in a variety of ways, those participating in this study agreed to accommodate for two distinct treatment groups: one in which students receive indirect, coded feedback, another that received direct corrections, and a third that received feedback on content and organization but not on grammatical accuracy. Feedback was administered digitally through GoogleDrive, where students were required to write all their assignments for the course. Digital administration of feedback by the researcher and research assistant allowed for more timely feedback to be provided and helped to manage the process and organize the data. Feedback was provided to students within 1-2 days of writing, which is timely compared to the reporting of WCF studies in general that average 2-5 days (Liu & Brown, 2015).

The researcher administered all the feedback for the essays that comprised the pre- and posttests (the first, second, seventh and eighth writings) for both treatment groups.

The direct feedback was administered using the “comment” feature in Google Docs by highlighting the erroneous portion of a sentence and then supplying the correct form without explanation in the margin. The “comment” feature was used rather than the “suggestion” feature for the direct feedback so that students could not simply click “accept changes” without paying attention to the feedback. With comments, unlike suggestions, students would have to read the comment to type the correction they see. The indirect coded feedback was administered by inserting the coded abbreviations immediately following an error in the text within brackets. The error and code in brackets were highlighted together in yellow (see Figure 1 for an example).

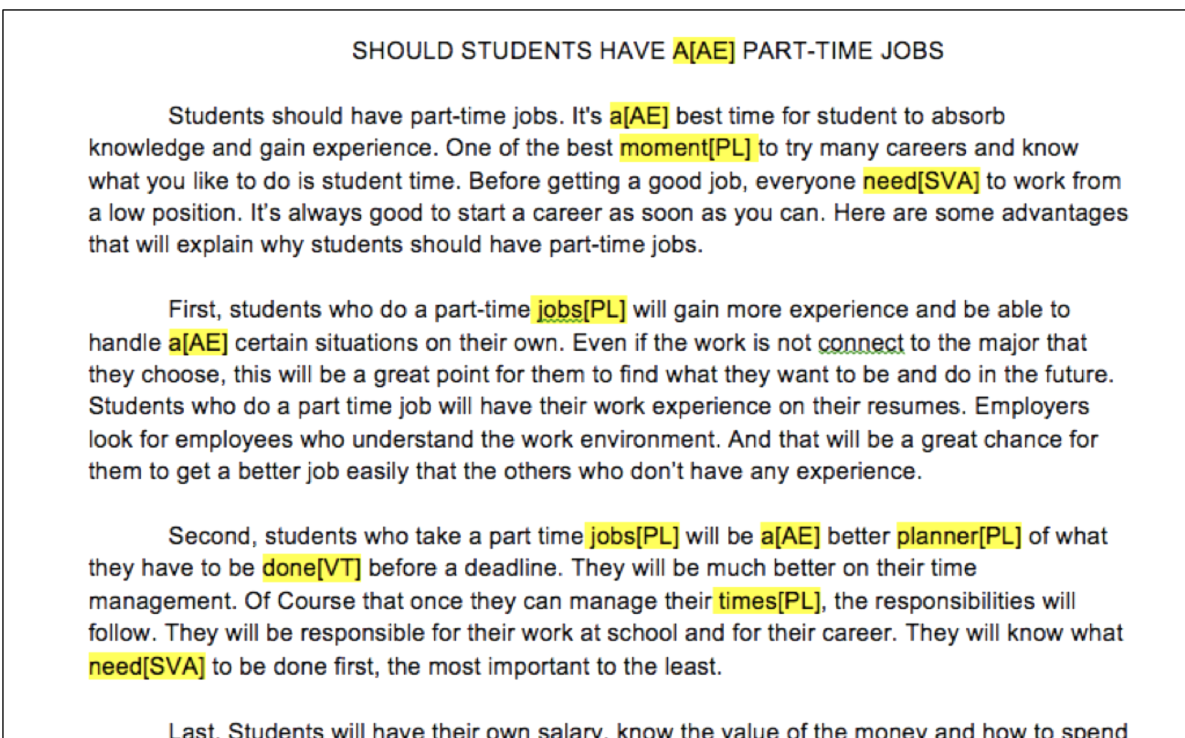


Figure 1. Example of indirect feedback treatment

## **Inter-Coder Reliability for Feedback Provision**

By the second week of each quarter, an applied linguistics graduate student with graduate level study in grammar assisted providing feedback on most of the treatment essays (the third through the sixth). Before the research assistant began providing feedback, training was conducted in two sessions. In the first session, the assistant was provided with examples of five essays that received coded feedback and five that received direct feedback and was instructed to review the examples of feedback, take notes, and then ask the researcher any questions that might arise. The assistant then coded another set of 10 writings, reaching an initial 83% inter-coder agreement. After analysis of the inter-rater coding, lower rates of agreement were found for verb tense / aspect (69%), singular-plural noun-agreement (81%), and articles (84%), while the other categories were high in agreement (above 90%). In the second training session, each case of verb tense / aspect and singular-plural agreement errors was discussed to find consistent discrepancies in coding decisions (e.g., interchangeable article and singular-plural errors, the inclusion of active / passive voice errors, which should have been excluded for that error type) and it was established that context in each essay would help to determine some of these distinctions for singular-plural noun agreement vs. article errors. After a second round of individual coding of an additional 10 essays from the pre-tests, inter-rater agreement improved to 93% and the assistant began helping with the treatment (essays 2-6). Inter-rater agreement was then measured between the researcher and assistant for an additional 20% of the data set that quarter (about 10% of the entire data set) at an acceptable 91.1% agreement.

The treatment lasted for five essays (over five weeks) and the second essay served as both the second half of the pre-test and the first treatment. Starting from the second essay, students' in each of the treatment groups began receiving feedback and were required to revise

their essays in response to the feedback before submitting a second draft that would be scored by the teacher. This arrangement controlled for revision (which has only been required in about 57% of WCF research; unreported in 23%, Liu & Brown, 2015), as students were provided with enough time to revise in the subsequent class meeting with opportunity to ask the teacher for help in revising. In all three groups, the teacher provided feedback on content and organization, as typically required in the curriculum. These global comments were provided on students' second drafts after accuracy was addressed in response to feedback on the first drafts.

### **Grammatical Knowledge Test development**

The Grammatical Knowledge Test was designed for the target population using authentic items procured from examples of errors in previous student writings in level 3 at the IEP. The test incorporated elements of previous tests of grammatical knowledge, in particular, previous metalinguistic knowledge measures. There were three sub-constructs of grammatical knowledge: (a) metalinguistic knowledge, (b) error correction ability, and (c) error type recognition. These sub-constructs represent the components of grammatical knowledge that together would help students to utilize feedback they receive (the indirect coded feedback, specifically, as it requires deeper engagement). This section will outline the test format, method of item development, piloting procedures, and item analysis and reliability.

**Test format.** The Grammatical Knowledge Test (presented in Appendix A) comprises three sections and 72 items in total. Appendix B outlines the test specifications. The first section combines the Metalinguistic Knowledge and Error Correction items for efficiency and represents the most common format previously employed to measure metalinguistic knowledge (refer to Table 4 in the literature review).

In the first section of the test, students are presented with authentic sentences from previous level 3 student writing that have an underlined grammatical error and students are asked to first make the correction by writing in the correct form (Error Correction), and then write the type of error followed by a description of the grammatical rule that is used to make the correction (Metalinguistic Knowledge). Instructions are presented along with the complete test in Appendix A, along with the rubric that was tailored for each item to score the metalinguistic knowledge items. The item format for the Metalinguistic Knowledge section follows Elder et al. (2007), which uses clearly distinct boxes for test-takers to both provide an error correction and explain the rule violation. The purpose of this distinction is to distinguish between a label for an error and the identification of a rule or pattern that governs the structure. Two examples are presented to begin the section, each on their own page to encourage test-takers to pay attention.

The second section includes 12 additional error correction items (the MK items were limited to 12 since they require the most time). The third section of the test included 36 multiple choice items that presented example sentences with underlined sections and asked test takers to select the error category that describes the underlined error presented. The underlined portions of the sentences varied from 3-4 words in length, which was adjusted in the piloting to improve item difficulty and/or discrimination. Distractors were also included that were not erroneous.

**A bottom-up approach to item validation.** As described in the literature review, a challenge and limitation of previous test designs of grammatical knowledge is the valid representation of the nuances present within linguistic structures. Even when a valid approach to selecting the representative linguistic structures is presented (such as the error analysis in this study), the challenge remains in deciding which examples best represent the construct. Considering that each linguistic structure (error type, in this study) represents varying degrees of

prototypical and peripheral usage of form-meaning relationships, more than a single item is necessary to support item validity for each error category. To determine the number of items and patterns of errors that represent the language use of the sample population, a “bottom-up” approach was used to analyze the sample of previous level 3 student writing. To determine the patterns within error types, the corpus of 40 essays was further analyzed to determine the variability required in designing items for each error type. This deeper analysis of the errors had a threefold goal; to determine (a) which error types require more than a single item to justify valid representation (i.e., would a single prototypical item suffice?), (b) how many different variations of items would reflect the range of the most relevant error patterns within each error type (i.e., the verb tense category encompasses more than simple past errors), and (3) exemplars of the most representative patterns (i.e., which types of errors in verb tense are most common for these Thai university students?).

In the analysis of the 20,392-word data set of previous writings, errors in the six target categories were further revisited for a more fine-grained analysis to code for patterns within each. Table 11 presents the results of the more fine-grained error analysis that determined the number and types of items to be included in the Grammatical Knowledge Test for each of the six error categories by providing the patterns, their frequency in the analysis, and accompanying authentic examples from the texts. Considering that the parameters of MK could vary drastically between error categories (e.g., the straight-forward rules associated with errors in subject-verb agreement compared to the range of patterns that comprise article usage), the aim here was to identify the most common error patterns in the target writing samples to guide the design of exemplar items. These exemplars represent the most typical patterns of errors for the most representative coverage in the test items. After coding the 605 errors to determine typical



patterns, a goal of 70% coverage was set within each error type. This was determined somewhat arbitrarily with the concern in mind to limit the total number of items in the metalinguistic knowledge section, as written explanations of rules are time-consuming and cognitively challenging for students. This goal was achieved by including one to three item types within each of the six error categories. Note that the different error pattern types in Table 11 are listed in

Table 11

*Common Error Patterns in Previous Level 3 Student Essays (40 essays, 4 different prompts)*

Distinct patterns within error types	Frequency of errors	Examples
<i>Singular-Plural Agreement</i> (1 item, 85% coverage)	193 total	
*Type 1. Missing –s on plural countable nouns	164 (85%)	-Many <u>school</u> need to find more teachers. -The students should study all <u>subject</u> in their first year.
Type 2. Unnecessary –s or –es ending to uncountable nouns	15 (7.8%)	-Children will gain <u>knowledges</u> from their teachers.
Type 3. Unnecessary –s or –es ending to singular nouns	14 (7.3%)	-Fear is the most important cause of this <u>problems</u> .
<i>Articles</i> (3 items, 70.5% coverage)	169 total	
*Type 1. Zero article when indefinite article necessary preceding pre-noun modifiers	52 (30.8%)	-That is <u>very difficult situation</u> for Thai people. -It is <u>better choice</u> for students in university.
*Type 2. Zero article when definite article necessary	40 (23.7%)	-This solution is <u>only way</u> to help them. -Math is part of <u>curriculum</u> that students are taught...
*Type 3. Unnecessary definite article with non-specific noun	27 (16%)	-This problem will cause Thailand to get into <u>the trouble</u> . -However, <u>the education</u> is needed to help society improve.

Type 4. Zero article when indefinite necessary with countable noun	20 (11.8%)	-Having <u>teacher</u> can encourage children to practice a lot.
Type 5. Unnecessary indefinite article with plural nouns	17 (10.1%)	-They will grow up to be <u>a</u> good adults in the future.
Type 6. Unnecessary definite article when indefinite necessary	13 (7.7%)	-Having a boyfriend at <u>the young age</u> can...
<hr/>		
<i>Subject-Verb Agreement</i> (2 items, 76.4% coverage)	84 total	
*Type 1. Missing <i>-s</i> or <i>-es</i> ending on verbs to agree with singular or collective/ uncountable subject	53 (62.1%)	-There is a program that <u>help</u> students to pay. -Education <u>encourage</u> students to...
*Type 2. Singular form of copular “be” with plural/collective noun	12 (14.3%)	-The security systems <u>is</u> strong enough. -People who <u>is</u> allowed more freedom can be happier.
Type 3. Plural form of copular “be” or “have” with singular noncount nouns	9 (10.7%)	-There <u>are</u> a lot of doubt about... -Every child <u>have</u> access to basic education.
Type 4. Unnecessary <i>-s</i> ending to verbs with modals or plural nouns	6 (7.1%)	-The solution will <u>works</u> better in Thailand. -Some parents <u>lets</u> their children study from home.
<hr/>		
<i>Part of Speech / Word Forms</i> (2 items, 79.7% coverage)	59 total	
*Type 1. Noun when adjective necessary	25 (42.4%)	-We need to find a <u>peace</u> solution to prevent war. -It helps if there is an <u>accident</u> case that is unplanned.
*Type 2. Adjective when noun necessary	22 (37.3%)	-People don’t know the <u>dangerous</u> of this problem. -It is better in terms of <u>religious</u> .
Type 3. Verb when Noun	12 (20.3%)	-But with the <u>utilize</u> of technology, it will...

necessary

---

<i>Verb Tense / Aspect</i> (3 items, 79% coverage)	**58 total	
*Type 1. Tense: past when present necessary	31 (53.4%)	-Teenagers nowadays <u>started</u> playing computer games at a very young age... -Over 800 million babies <u>were born</u> every year.
*Type 2. Aspect: attempt at progressive when simple is more natural (or when correct form of progressive would work)	8 (13.8%)	-Thus, when these children <u>growing up</u> , they will become hard working. -Every morning people in Thailand <u>are facing</u> traffic jams.
*Type 3. Aspect: simple when perfect appropriate	7 (12.1%)	-In the last ten years, climate change <u>becomes</u> the most significant issue around the world. -Since 2010, over a hundred schools <u>join</u> this program.
Type 4. Aspect: incorrect construction of progressive	5 (8.6%)	-We have been trying this solution for three years but we see it <u>is fail</u> right now so we need to change it.
<hr/>		
<i>Run-On Sentences</i> (1 item, 89.2% coverage)	37 total	
*Type 1. Two independent clauses without appropriate subordinator/ conjunction / punctuation	33 (89.2%)	-It teaches students everything about <u>business it</u> emphasizes how to start your own business. -No one is too young to decide what they <u>want</u> , <u>this</u> is also true for children too.
Type 2. Too many coordinators/ adverbials	4 (10.8%)	-Every class might contain about 40 kids, however there is only one male, so it is not even, so mostly women study this subject.

---

\*Item types chosen for inclusion in the test represent each of the categories with an asterisk.

\*\*Not all error types are listed as a sub-category could not be identified in a few instances.

order of frequency (within each of the six error categories) and the top selections in each

category were chosen for inclusion until reaching the desired coverage of 70% within each error

category. The pattern types included in the Grammatical Knowledge Test are marked by asterisks.

Once the 12 patterns were identified within the six broad error categories, a list of 12 error patterns was consulted with the help of another researcher in applied linguistics who is a native Thai speaker and an experienced English teacher at the university level in Thailand. Together, we selected several examples of each error type in the previous student writings that seemed clear and representative of the patterns of error without being ambiguous as to alternative interpretations. This analysis resulted in the complete set of test items used in each of the three sections of the test. Table 12 presents each item on the test organized by error category and error pattern type within each category organized by section (i.e., Metalinguistic Knowledge, Error Correction, and Error Type Recognition) and corresponding item numbers. Minor edits were made in some of these authentic errors to shorten the length of sentences and to clarify the context efficiently (e.g., adding an adverbial to make the time period clear for a verb tense error).

Table 12

*Test Items (organized by item type within error categories)*

Error Category	Type within Category	Items (MK = Metalinguistic Knowledge, EC = Error Correction, ETR = Error Type Recognition multiple choice)
SP (1 type)	Missing inflectional –s on plural countable nouns	MK#6, EC1#6- Many <u>school</u> need to find more teachers. EC2#1- The students should study all <u>subject</u> in their first year. ETR1#1- These types <u>of student will</u> learn faster. ETR2#30- Many <u>parent</u> can believe in the quality of the education here. ETR3#38- We can find it in media such as television, <u>magazine</u> , and newspapers.
Art (3 types)	Type 1: Zero article when indefinite appropriate preceding pre-noun modifiers	MK#8, EC1#8- That <u>is very</u> difficult situation for Thai people. EC2#9- It is safe at home so it <u>is much</u> better environment for kids to study. ETR1#2- Parents can indicate what <u>is good</u> thing to do for

		<p>their children.</p> <p>ETR2#8- Most parents want their kids to <u>have better</u> future.</p> <p>ETR3#11- I hope to <u>earn good</u> grade in this class.</p>
	Type 2: Zero article when definite article appropriate	<p>MK#12, EC1#12- Some classes have better teachers. For example, in math class teacher is teaching well.</p> <p>EC2#10- That type of school <u>has worst</u> environment for kids.</p> <p>ETR1#14- Some people believe <u>that homeschooling</u> trend is negative.</p> <p>ETR2#18- The first subject is at 8am <u>and last</u> subject ends at 3pm.</p> <p>ETR3#25- You can avoid a traffic jam by <u>using sky</u> train in Bangkok.</p>
	Type 3: Unnecessary definite article with non-specific noun and no prior reference	<p>MK#4, EC1#4- This problem will cause Thailand to get <u>into the trouble</u>.</p> <p>EC2#7- Students need to work hard and take <u>the time</u> every day to study.</p> <p>ETR1#27- Students of all ages need to <u>do the homework</u>.</p> <p>ETR2#10- This will <u>the</u> big problem if we do not solve it now.</p> <p>ETR3#35- This issue is a useful thing to talk about in <u>the society</u> in every country.</p>
SV (2 types)	Type 1: Missing inflectional –s ending on verbs to agree with singular or collective/ uncountable noun	<p>MK#2, EC1#2- The program <u>help</u> students to pay for their books.</p> <p>EC2#4- Everything that <u>appear</u> in the homework will be on the test.</p> <p>ETR1#3- Today the debate <u>begin</u> about the best solution.</p> <p>ETR2#7- Therefore, my school <u>encourage</u> students to succeed in their lives.</p> <p>ETR3#20- This method just <u>tell</u> them it is wrong without the reason why.</p>
	Type 2: Singular form of a copular “be” with plural/collective noun	<p>MK#9, EC1#9- The two security systems we have <u>is</u> strong enough.</p> <p>EC2#8- Some people believe those girls who have a baby before they get married <u>is</u> bad people.</p> <p>ETR1#26- When children <u>is</u> restricted by their parents they will grow up slower than children with freedom.</p> <p>ETR2#32- If people <u>is</u> allowed more choices they will be happier.</p> <p>ETR3#40- Some parents choose to educate their children at home because there <u>is</u> so many benefits.</p>
POS (2 types)	Type 1: Noun when adjective appropriate	<p>MK#5, EC1#5- We need to find a <u>peace</u> solution.</p> <p>EC2#3- It is not <u>safety</u> for students to stay out late at night.</p> <p>ETR1#6- We need a <u>logic</u> solution to this simple problem.</p> <p>ETR2#29- Children will not have the same <u>society</u> experience if they stay at home.</p>

		ETR3#17- If someone gets <u>pregnancy</u> as a teenager it could cause difficulty in their family.
	Type 2: Adjective when noun appropriate	MK#10, EC1#10- People don't know the <u>dangerous</u> of this problem. EC2#12- Students may receive several <u>beneficial</u> from studying at university. ETR1#22- It won't happen again because that was an <u>accident</u> case. ETR2#13- Therefore, teaching the <u>true</u> is the best thing parents can do for their children. ETR3#37- It is better to provide children many <u>optional</u> and let them make their own choice.
VT/A (3 types)	Type 1: Tense: past when present appropriate	MK#1, EC1#1- Currently, over 100 million babies <u>were born</u> every year. EC2#2- Therefore, today I strongly <u>believed</u> that parents should be strict to their children. ETR1#4- If they <u>failed</u> when they are older, they will find a solution by themselves. ETR2#34- Lots of things are changing every day as more technology <u>was invented</u> all the time. ETR3#16- Parents are important for children when they are young because they <u>were in control</u> of their children.
	Type 2: Aspect: erroneous construction of progressive (with modal error or when simple appropriate)	MK#11, EC1#11- Thus, when these children <u>growing up</u> , they will become responsible. EC2#5- They will be worried to ask and will not <u>talking</u> with their parents about the problem. ETR1#12- Students still <u>do not understanding</u> about the issue properly. ETR2#19- However, as <u>the world moving</u> forward, we need more solutions. ETR3#21- We have been trying this solution for three years but every <u>time it failing</u> .
	Type 3: Aspect: simple when perfect or past appropriate	MK#3, EC1#3- In the last decade, climate change <u>becomes</u> the most significant issue around the world. EC2#11- Since 2010, over one hundred schools <u>join</u> the program. ETR1#28- Since I was 10 years old, I <u>grow</u> 60 cm. ETR2#3- Since last year <u>this become</u> a very common problem. ETR3#39- I <u>never sing</u> that song in my life.
RO (1 type)	Two independent clauses without appropriate subordinator/ conj/ punct	MK#7, EC1#7- It teaches everything about <u>business, it</u> emphasizes how to start your own business. EC2#6- Parents should be role models for their <u>children, they</u> will learn more positive things from their parents. ETR1#15- No one is too young to decide what they <u>want, this</u> is also true for children too.

---

ETR2#24- Children will gain a lot of knowledge, maybe they can learn from friends or from traveling too.

ETR3#36- If parents give their children too much freedom, they will risk having bad behavior, it can be true in some cases.

---

**Piloting of the test.** The Grammatical Knowledge test was first piloted with five graduate students in applied linguistics (native and non-native English speakers), and following revision, was then piloted twice in the target context in Thailand in two level 3 classes ( $n = 44$ ), first in January and then in February of 2016. Two sets of item analyses led to substantial revision and item replacement to improve item difficulty and discrimination, as well as test reliability. Through piloting, the scoring system for the Metalinguistic Knowledge items was revised to incorporate distinct criteria for each item, also following Elder et al. (2007), who made available their complete scoring key on the IRIS database (the digital repository of data collection instruments for research into second language learning and teaching). As presented above in Table 6, their scoring criteria consisted of 3 levels (0, 1, and 2 points) with descriptors tailored to each error type. Of the ten targeted error types in Elder et al. (2007), five overlapped with those in the present study, which were adapted for the items created from authentic errors from the Thai students' writings. The final test results are presented in the results section with item analysis and reliability.

### **Data Collection Procedures**

Data collection occurred in the first seven weeks of two consecutive academic quarters in three writing class sections. Data was collected on GoogleDrive, where students completed all their writings in class and completed the Grammatical Knowledge Test online. This section will outline the sequencing of the data collection in terms of the administration of the Grammatical Knowledge Test, the training procedures and materials used to prepare students in the indirect

coded feedback group to understand how to make use of the feedback they received, and the data collection procedures for the writing tasks.

**Test administration.** In the first week of the semester (the second class meeting) the Grammatical Knowledge Test was administered to all groups using Survey Gizmo. A counterbalanced design was used so that half the students in each group took one of two versions with items presented in opposite order to control for test fatigue. The test administration was in-class and required between 40-60 minutes for each class to complete. The teachers were instructed to guide students through the instructions and to check answers to the practice items as a class before students began the test individually. The teachers were also instructed to remind students to work individually, as the results of the test could not hurt their grades, but rather could help them to identify the areas in which they needed the most help. Students were not provided with scores of their Grammatical Knowledge Test until week eight of the course, following the data collection of the posttest. At that time, students were provided with answers to the test with a score report that separated Grammatical Knowledge scores for each of the six error types to illustrate individual strengths and weaknesses, along with corresponding pages in the course book that covered exercises for these grammar points explicitly. This Grammatical Knowledge test, in an abbreviated form that does not include the metalinguistic knowledge section, has been adopted into the ongoing IEP level 3 curriculum as a diagnostic that can be automatically scored at the start of the course.

**Training students to use the indirect coded feedback.** Students in the indirect feedback group participated in a 30-minute training activity to be introduced and acclimated to the coded feedback system and learn how to respond to it successfully immediately prior to receiving the first round of feedback. Previous research has highlighted the importance of



training students to make use of codes (e.g., Ferris, 2011; Brown, 2012), although most WCF studies do not report whether training is provided (Liu & Brown, 2015). The training session was conducted using materials presented in Appendix C, which included two structured activities for practice. Part 1 of the training began with an exercise that presented isolated sentences containing errors that represent the six linguistic targets. In pairs, students first attempted to make corrections and were then guided to refer to the list on the same page that presented the meaning of the codes. At this point, students were also introduced to the Grammar Code Key (Appendix D) that presents the different error categories and abbreviations and presents additional examples of each error type with suggestions to help in making revisions. While working on this activity, students were encouraged to ask their teacher questions and answers were reviewed and discussed as a class before moving to the second activity in the training.

Part 2 of the training activity required students to log into a sample student writing on Google Doc (see Appendix C, part 2) on their computers in class. This sample student essay (an introduction paragraph to an essay) was adapted from a previous level 3 student essay and included errors in each of the six categories to represent the most common patterns of errors within each category. In part 2, students were instructed to work individually to follow the comments previously provided by a teacher on screen, simulating the process of responding to feedback they would experience throughout the semester using the coded symbols (e.g., SV, SP) provided. Their answers were checked on GoogleDrive to ensure that they understood the coded feedback and could use it successfully in revising so that instructors could address confusion or questions at the onset of the treatment. In addition, the researcher or research assistant pasted the Grammar Code Key to the top of every marked essay that students received back to ensure students had visible access to the code for reference when making revisions on each writing.

**Writing data collection.** The writing data was collected over the first seven weeks of the 10-week quarter to allow for the final three weeks to be devoted to a focus on grammatical feedback for the control group that was withheld for the duration of the study. As Figure 2 illustrates, the pretest combined two in-class writings from the first week of the semester. Although most WCF studies have used only a single writing assignment as a pre- or posttest, typically just 100-300 words in length (Liu & Brown, 2015), such limited writing length raises concerns of reliability in representing students' performance in accuracy. A single page of writing seems unconvincing in its representation of accuracy at a given point in time, especially

<b>Group Condition</b>	<b>Grammar Test</b>	<b>Pretest</b>	<b>Treatment/ Control Period</b>	<b>Posttest</b>
Direct CF <i>n</i> = 41	Week1	Week 1	Weeks 2, 3, 4, 5, 6	Week 7
Indirect CF <i>n</i> = 44	Grammatical Knowledge Test administered	In-class writings #1 & 2 ( <i>M</i> = 833 words)	5 treatments on in-class writing assignments  ( <i>M</i> = 410 words per treatment)	In-class writings #7 & 8 combined  ( <i>M</i> = 808 words)
Control (no grammar CF) <i>n</i> = 42				

Figure 2. Data collection procedures.

considering the multitude of factors that likely influence performance in written accuracy. When writing topic varies between the pre- and posttests, as is the case necessarily in most classroom contexts, combining two separate (but consecutive) data points for each pre- and posttest may help limit the influence of a single topic that could cause deviation in performance. This rationale, along with the relatively high frequency of writings in this course allowed for a lengthier pre- and posttest for more reliable indication of student performance while still

maintaining a generous supply of five treatments, which represents the median number of treatments in WCF studies that include more than one treatment (Liu & Brown, 2015).

It should be noted that each of the sections incorporated peer-review to a minor extent (on one to three essays, as reported by teachers) in the revision process, but teachers were instructed to limit peer-provided feedback to content and organization. It is likely that not all students avoided providing feedback on grammar in these sessions, and this could not be controlled systematically. However, the emphasis on other areas for peers to review (with the use of guided peer-review worksheets) lends confidence that peer feedback on grammar had limited influence as a potential confounding variable.

### **Analysis**

The initial data analysis involved screening to identify participants with missing data. Participants were excluded from the study if they were missing one of the four essays that served as the pre- and posttest data and/or the Grammatical Knowledge Test. When students were absent for a writing, they were encouraged by their teachers to complete the writings in the following class to limit attrition. Only two students in the study (who completed the course) were excluded because they could not complete all the pre- or posttests. Attendance was high in the program, although students occasionally missed an in-class writing during the treatment and were then required to make-up the writings they missed for homework. Similarly, the few students absent on the day of the Grammatical Knowledge test took the test in the following class meeting. These procedures limited total attrition to eight students, four of whom did not complete the course.

**Linguistic measures of the dependent variable.** Each of the pre- and posttest texts were analyzed for linguistic accuracy in the six targeted error types. Following previous studies

exploring the effectiveness of WCF (e.g., Chandler, 2003; Truscott & Hsu, 2008; van Beuningen et al., 2012), an error ratio was used to measure accuracy that was normed, i.e., (number of linguistic errors/total number of words) x 1,000. A 1,000-word ratio rather than the more common 100-word ratio was used because the pre- and posttests were closer to 1,000 words each ( $M = 820$  words). Once the normed accuracy measure was calculated for the pre- and posttests, a gain score was determined by subtracting the normed accuracy rates of the posttests from the pretests for each error type. The normed gain scores served as the measure of the dependent variable for each group. To investigate the interaction between gain scores (the dependent variable) and grammatical knowledge (research questions 3 and 4),  $Z$ -scores were calculated for both variables for statistical analysis.

**Statistical analyses.** Two sets of data were entered from Excel into IBM SPSS Statistics 24 for analysis. The first set were the normed accuracy measures for each writing task and combined to form the pre- and posttests. The second set of data were the Grammatical Knowledge Test scores for each sub-section and the cumulative total scores. To compare the outcomes of the different feedback types relative to the control (research questions 1 and 2), descriptive statistics including mean gain scores in accuracy from pre- and posttests, standard deviations, effect sizes, and confidence intervals were calculated. Following an assumptions check, a one-way ANOVA was used to compare overall gain scores between the three groups. The statistical significance level was set at .05, the standard in educational research, while the effect sizes were calculated and reported using Cohen's  $d$  for comparison of mean differences. To interpret the magnitude of the effect sizes, Plonsky and Oswald's (2014) field-specific benchmarks are referenced ( $d = .4$  as small,  $d = .7$  as medium,  $d = 1.0$  as large), rather than Cohen's traditional interpretations ( $d = .2$  for small,  $d = .5$  for medium, and  $d = .7$  for large). It

should be noted though, that eventually these field-specific benchmarks for interpreting effect sizes will likely be further specified according the aggregate comparable results in sub-domains of applied linguistics, and Kang and Han's (2015) meta-analysis results of effect sizes for gains in accuracy ( $n = 22$ ) should be consulted when interpreting results for WCF studies moving forward.

Research Question 3 investigates the independent variable of grammatical knowledge as moderator of effectiveness. Pearson correlation was used to explore potential relationships between grammatical knowledge and gains in accuracy between groups and within the full sample. Grammatical knowledge comprised an aggregate score for all three sub-constructs tested, and meta-linguistic knowledge was also correlated with gains in accuracy in isolation. For both Research Questions 3 and 4 multiple regression (MR) was the planned statistical procedure had the analysis of correlations revealed significant patterns. MR could reveal the interaction effect between feedback type and grammatical knowledge within error types, following previous ATI studies in SLA (e.g., Li, 2013; Yilmaz, 2013), which have focused on the interaction between instructional treatment types and individual learner aptitude differences. As Li (2013, p. 644) discusses, either MR or analysis of covariance (ANCOVA) could be suitable statistical techniques to answer questions of interaction between multiple variables, depending on the substantive focus. MR was deemed a suitable statistic because it allows for the inclusion of multiple continuous variables (i.e. gain scores and grammatical knowledge) along with a categorical variable (i.e., feedback type) if a dummy variable is computed. MR could be used to examine how multiple IVs (i.e., grammatical knowledge and feedback type) individually and collectively predict the development in accuracy in response to WCF in writing. Once the variance in gain scores is determined for grammatical knowledge, then the treatment type can be

entered into the equation (hierarchically) to determine the interaction effect of grammar knowledge and treatment type. Rather than rely on a bivariate correlation on two different sets of data, MR is the multivariate equivalent that could explain the variance in the gain scores as a function of both of the independent variables—grammatical knowledge and feedback type (Jeon, 2015; Larson-Hall, 2009). MR would also enable prediction of the value of the type of feedback with respect to grammatical knowledge. An advantage of MR over ANCOVA here is that ANCOVA requires categorical independent variables, but since this study involves a continuous dependent variable (gain scores), one continuous independent variable (grammatical knowledge), and another categorical independent variable (feedback type), the use of ANCOVA would require the continuous grammatical knowledge scores to be divided into levels (i.e., high, mid, low), which is not advisable due to loss of variance (Plonsky, 2015).

In hierarchical MR, Grammar Knowledge scores could be entered into the equation first, followed by feedback type to determine the contribution of feedback type once grammar knowledge is accounted for (Jeon, 2015; Larson-Hall, 2009). The suitability of MR in this study would only be to examine the interaction between grammatical knowledge, feedback type, and gains in accuracy for the full sample ( $N = 127$ ) as the number of variables and statistical tests require significant power for confidence in findings, and testing within each error type would lower power substantially. A Pearson correlation matrix (grammatical knowledge Z-scores x Z-score gains in accuracy) was first used to determine if a relationship existed between grammatical knowledge and gains in accuracy in general and across each linguistic feature.

## **Chapter 4**

### **Results**

This study investigated four main issues: (a) the relative effectiveness of direct and indirect WCF in developing learners' accuracy, (b) the relative amenability of different linguistic targets to WCF in developing learners' accuracy in response to the different feedback types, (c) the role of grammatical knowledge in moderating the effectiveness of WCF in developing grammatical accuracy, and if grammatical knowledge is found to play a role, (d) the potential interaction between grammatical knowledge, feedback type, and gains in accuracy for individual linguistic structures. This chapter presents the results of the study organized by the research questions. First, normed error rates across each writing task are presented, followed by results of an equivalency check of pretest accuracy between groups. Analyses regarding treatment effects are presented by overall gains in accuracy between groups and by individual target structures. Grammatical Knowledge Test scores are then presented (along with results for each section of the test) with item analysis and test reliability measures. To determine if Grammatical Knowledge co-occurs with gains in accuracy, correlations are analyzed across groups. A multiple regression model is not required, as no substantial patterns of correlation exist at the group level (a prerequisite for utilizing multiple regression). Finally, justification for conducting a post hoc analysis and the methods employed are described with results presented of analysis of data by comparing groups made up of the extreme cases of grammatical knowledge.

#### **Comparability of Groups at the Onset of the Study**

The comparability of students' overall writing proficiency was presented in the methods section (Participants and Grouping), showing that the placement of students into PC level 3 is valid according to IELTS scoring of the placement tests for writing. Additionally, normed

accuracy rates on the initial writing assignment served as a baseline to determine equivalency between groups at the outset of the study. Table 13 presents the descriptive statistics for the comparability of overall accuracy (and by error type) between groups on the baseline essay. Note that higher mean scores represent a greater number of errors.

Table 13

*Equivalency between Groups for Pretest Accuracy (Normed to 1,000 words)*

CF Type		*SP	Art	SV	POS	VT	RO	Total
Direct	<i>M</i>	20.68	15.62	6.16	4.39	5.96	3.18	55.98
	<i>SD</i>	11.23	8.81	4.39	2.94	3.66	3.99	22.22
Indirect	<i>M</i>	21.32	16.50	6.06	4.11	5.55	1.70	55.24
	<i>SD</i>	9.64	8.00	3.76	3.04	3.84	1.78	17.84
Control	<i>M</i>	19.62	18.20	5.64	5.55	5.65	2.60	57.26
	<i>SD</i>	9.58	10.07	3.58	3.17	5.24	3.10	21.58
Total	<i>M</i>	20.55	16.78	5.95	4.68	5.72	2.47	56.15
	<i>SD</i>	10.01	8.98	3.89	3.09	4.27	3.11	20.44

\*SP = singular-plural agreement, Art = articles, SV = subject-verb agreement, POS = part of speech (word form), VT = verb tense/aspect, RO = run-on sentences

A one-way analysis of variance (ANOVA) calculated on students' baseline accuracy with group condition as a between-subjects variable showed that there were no significant differences among the three groups in overall accuracy on the first writing task,  $F(2,124) = .11$ ,  $p = .90$ ,  $\eta^2 = .002$ . These findings ensure that the three groups were comparable at the onset of the study. Table 14 presents the complete data set for normed accuracy measures between groups and error types across each individual essay that comprised the pre- and posttests (essays 1, 2, 7, and 8).



The total accuracy rate in this table reveals overall equivalency between the first ( $M = 56.07$ ) and second essays ( $M = 56.09$ ) as well.

Table 14

*Descriptive Statistics of Accuracy by Individual Essay*

WCF Type	Essay #		SP	Art	SV	POS	VT	RO	Total
Direct	1	<i>M</i>	21.83	15.13	6.20	4.24	6.40	3.12	56.92
		<i>SD</i>	14.44	8.76	6.35	3.53	4.38	4.34	24.94
	2	<i>M</i>	19.05	16.05	5.91	4.56	5.46	3.23	54.26
		<i>SD</i>	11.43	10.74	4.75	3.88	4.90	4.32	24.30
	7	<i>M</i>	13.46	13.64	6.35	4.87	4.95	4.87	44.58
		<i>SD</i>	10.87	8.58	4.30	4.32	5.31	4.32	21.05
	8	<i>M</i>	21.49	14.74	7.76	4.14	4.71	2.26	55.10
		<i>SD</i>	14.41	11.14	6.09	4.14	4.57	3.64	26.47
Indirect	1	<i>M</i>	22.00	13.67	6.50	4.03	4.68	2.02	52.90
		<i>SD</i>	14.29	8.18	4.63	4.37	3.74	2.67	21.83
	2	<i>M</i>	20.89	19.33	5.61	4.17	6.46	1.32	57.77
		<i>SD</i>	8.77	11.93	4.27	3.71	5.54	1.90	22.16
	7	<i>M</i>	13.43	16.95	7.19	5.13	4.48	5.13	49.33
		<i>SD</i>	9.56	11.03	6.01	4.56	4.19	4.56	23.29
	8	<i>M</i>	20.98	11.65	4.15	3.19	3.78	0.97	44.71
		<i>SD</i>	15.95	7.45	3.88	3.24	5.43	1.96	24.26
Control	1	<i>M</i>	20.30	17.62	5.58	5.64	6.37	3.05	58.56
		<i>SD</i>	14.62	11.93	4.76	4.36	7.05	3.91	25.83
	2	<i>M</i>	19.18	18.68	5.62	5.58	4.85	2.22	56.13
		<i>SD</i>	10.09	11.12	4.01	3.79	4.56	2.71	22.35
	7	<i>M</i>	18.58	19.94	7.85	6.66	5.62	6.66	60.81
		<i>SD</i>	11.55	10.23	6.53	4.23	4.75	4.23	22.69
	8	<i>M</i>	22.62	17.46	5.87	5.29	4.45	1.90	57.58

		<i>SD</i>	14.23	10.08	4.92	4.36	4.39	2.34	23.93
Total	1	<i>M</i>	21.38	15.45	6.10	4.63	5.80	2.71	56.07
		<i>SD</i>	14.35	9.81	5.26	4.15	5.26	3.70	24.14
	2	<i>M</i>	19.73	18.06	5.71	4.76	5.60	2.23	56.09
		<i>SD</i>	10.08	11.29	4.32	3.81	5.03	3.19	22.80
	7	<i>M</i>	16.65	16.87	7.51	5.55	5.01	5.55	51.59
		<i>SD</i>	11.08	10.27	5.98	4.41	4.74	4.41	23.22
	8	<i>M</i>	22.77	14.57	6.24	4.21	4.28	1.69	52.32
		<i>SD</i>	15.17	9.84	5.51	3.98	4.81	2.76	25.33

### Results of Research Question 1: Relative Effectiveness of WCF Types

The dependent variable (gain scores in accuracy) was calculated by subtracting the normed accuracy of writings 7 and 8 (combined to serve as the posttest) from the normed accuracy of writings 1 and 2 (combined as the pretest). Note that positive values indicate development, i.e., a gain in the normed accuracy rate from the pretest to the posttest. The first research question investigates the relative effectiveness of the different feedback types on overall accuracy in comparison to a control group. Descriptive results show that the gains for the direct feedback group ( $M = 5.96$ ,  $SD = 15.53$ , Cohen's  $d = .28$ , 95% CI [1.06, 10.86]) and for the indirect feedback group ( $M = 8.33$ ,  $SD = 16.78$ , Cohen's  $d = .43$ , 95% CI [3.23, 13.43]) were greater than that of the control group ( $M = -1.70$ ,  $SD = 19.55$ , Cohen's  $d = -.08$ , 95% CI [-7.80, 4.39]). A one-way ANOVA was also calculated after testing for assumptions required for the analysis of variance. All the assumptions were met in that there were no extreme outliers in the data (more than +/- 3 standard deviations) as assessed by inspection of boxplots, the CWWS score was normally distributed in each group, as assessed by Shapiro-Wilk's test ( $p = .22$ ,  $.34$ ,  $.70$ ), and there was homogeneity of variances, as assessed by Levene's test for equality of

variances ( $p = .46$ ). In addition, descriptive statistics were compared to explore potential instructor effect, as the instructors for the direct and indirect groups switched between the two quarters that data were collected for a counter-balanced design. Exploratory analysis suggests limited effect for instructor, as the direct group in first quarter (teacher A:  $n = 18$ ,  $M = 4.81$ ,  $SD = 13.63$ ) was slightly outperformed in terms of gains in accuracy by the direct group in the second quarter (teacher B:  $n = 23$ ,  $M = 7.18$ ,  $SD = 16.90$ ), but a similar opposite difference could be found between the indirect groups comparing the first quarter (teacher B:  $n = 21$ ,  $M = 6.72$ ,  $SD = 17.83$ ) and the second quarter (teacher A:  $n = 23$ ,  $M = 10.02$ ,  $SD = 15.18$ ). These differences were comparable to the two control groups with the same teacher each quarter (first quarter:  $n = 22$ ,  $M = .46$ ,  $SD = 17.21$ ; second quarter:  $n = 20$ ,  $M = -3.90$ ,  $SD = 21.81$ ).

Results of the one-way ANOVA show that the gains in accuracy were significantly different between the groups,  $F(2, 124) = 3.88$ ,  $p = .023$ ,  $\eta^2 = .07$ . A Tukey post hoc pairwise comparisons revealed that gains in accuracy for each of the treatment groups (direct and indirect) were significantly greater than that of the control group,  $p = .047$ ,  $p = .008$ , respectively. However, no statistical difference was found for gains in accuracy between the direct and indirect treatment groups ( $p = 0.532$ ), indicating that both feedback groups outperformed the control group in developing accuracy between the pre- and posttests, but that one feedback group did not significantly outperform the other. Standard error, confidence intervals, and minimum and maximum scores are presented in Table 15.

Interpreting these effect sizes in view of Plonsky and Oswald's (2014) field specific benchmarks, and specifically for within-group study designs (i.e., 25<sup>th</sup> percentile  $d$  values of .60 considered small, 50<sup>th</sup> percentile of 1.00 as medium, and 75<sup>th</sup> percentile of 1.40 as large), the

Table 15

*Comparison of Overall Gains in Accuracy between Groups*

	<i>N</i>	<i>M</i>	<i>SD</i>	Std. Error	95% CI		Minimum	Maximum
					Lower	Upper		
Direct	41	5.96	15.53	2.42	1.06	10.86	-22.68	53.68
Indirect	44	8.33	16.78	2.53	3.23	13.43	-39.56	42.48
Control	42	-1.70	19.55	3.02	-7.80	4.39	-39.27	45.30
Total	127	4.25	17.76	1.58	1.13	7.37	-39.56	53.68

indirect and direct feedback resulted in relatively small improvement in grammatical accuracy after the five treatments, while the control group experienced no statistically significant gain. Figure 3 is a visual representation of the outcomes for differences between mean gains in overall accuracy between the three groups, with error bars showing 95% confidence levels.

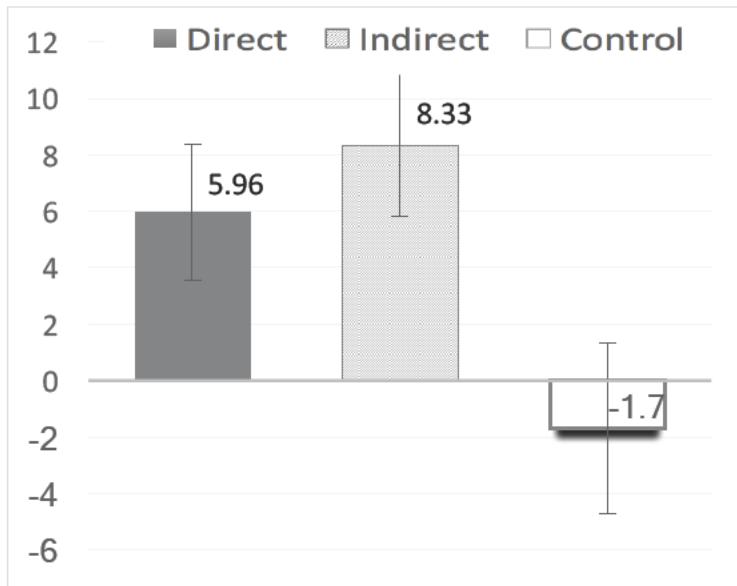


Figure 3. Overall gains in accuracy by treatment type.

A surprising pattern in the results reveals that the control group generally decreased in accuracy from the pre- to the posttest measures, although not at a significant level. Referencing Table 14 above, overall accuracy for writing number seven ( $M = 51.59$ ,  $SD = 23.22$ ) was slightly lower than that of writing eight ( $M = 52.32$ ,  $SD = 25.33$ ), relative to nearly no difference between the pre-test writings one ( $M = 56.07$ ,  $SD = 24.14$ ) and two ( $M = 56.09$ ,  $SD = 22.80$ ), suggesting that writing seven could have been slightly more difficult than the other writing topics that comprised the pre- and posttest. In retrospect, the topic for writing seven, “*Should the use of pesticides and/or GMO’s be banned in Thailand?*” seems more difficult than the others, which likely relate more directly to the students’ lives (e.g., “*Should children be allowed to use mobile phones?*”). Considering the timed and unreferenced conditions of these writing assignments, lack of familiarity with a topic could play a confounding role in the accuracy outcome. It should be noted that the standard deviation for the gains of each of the three groups are quite high (15.53, 16.78, 19.55), generally two to three times the mean, suggesting unusually high variation in performance.

## **Results of Research Question 2: Relationship between WCF Type and Error Type**

The second research question concerns the effectiveness of the different WCF types relative to the linguistic targets. Table 16 presents the descriptive statistics (mean gain scores, standard deviation, effect size as Cohen’s  $d$ , and 95% confidence intervals) for accuracy by individual error type and for overall accuracy across the three groups. Figure 4 presents a visual representation of the relative amenability to WCF for each of the feedback types (as compared to the control group) for each of the six target error types, along with standard error bars of 95% confidence. As illustrated in the Figure 4, the error type for which the largest gains (represented in the y-axis) resulted for either treatment was for singular-plural noun agreement, where gains

Table 16

*Gains in Accuracy between Pre- and Posttests (positive values indicate increase in accuracy)*

		S/P	Art	S/V	POS	VT	RO	Overall
Direct ( <i>n</i> = 41)	M	3.19	1.34	-.91	-0.14	1.10	1.38	5.96
	SD	10.22	8.72	5.45	3.27	4.35	3.39	15.53
	<i>d</i>	.28	.16	-.22	-.05	.29	.41	.28
	CI (upper)	6.42	4.10	.81	.90	2.47	2.45	10.86
	CI (lower)	-.03	-1.41	-2.63	-1.17	-.28	.31	1.06
Indirect ( <i>n</i> = 44)	M	4.15	2.27	.41	-0.02	1.39	0.12	8.33
	SD	10.84	9.92	5.28	4.09	4.58	2.70	16.78
	<i>d</i>	.40	.29	.10	-.01	.36	.06	.43
	CI (upper)	7.45	5.29	2.02	1.22	2.78	.94	13.43
	CI (lower)	.86	-.74	-1.19	-1.27	.00	-.70	3.23
Control ( <i>n</i> = 42)	M	-.80	-.49	-1.23	-.38	0.59	0.59	-1.70
	SD	10.66	10.57	3.90	2.99	5.43	2.11	19.55
	<i>d</i>	.08	-.05	.29	-.12	.13	.23	-.08
	CI (upper)	2.52	2.81	-.01	.55	2.28	1.25	4.39
	CI (lower)	-4.12	-3.78	-2.44	-1.31	-1.10	-.07	-7.80

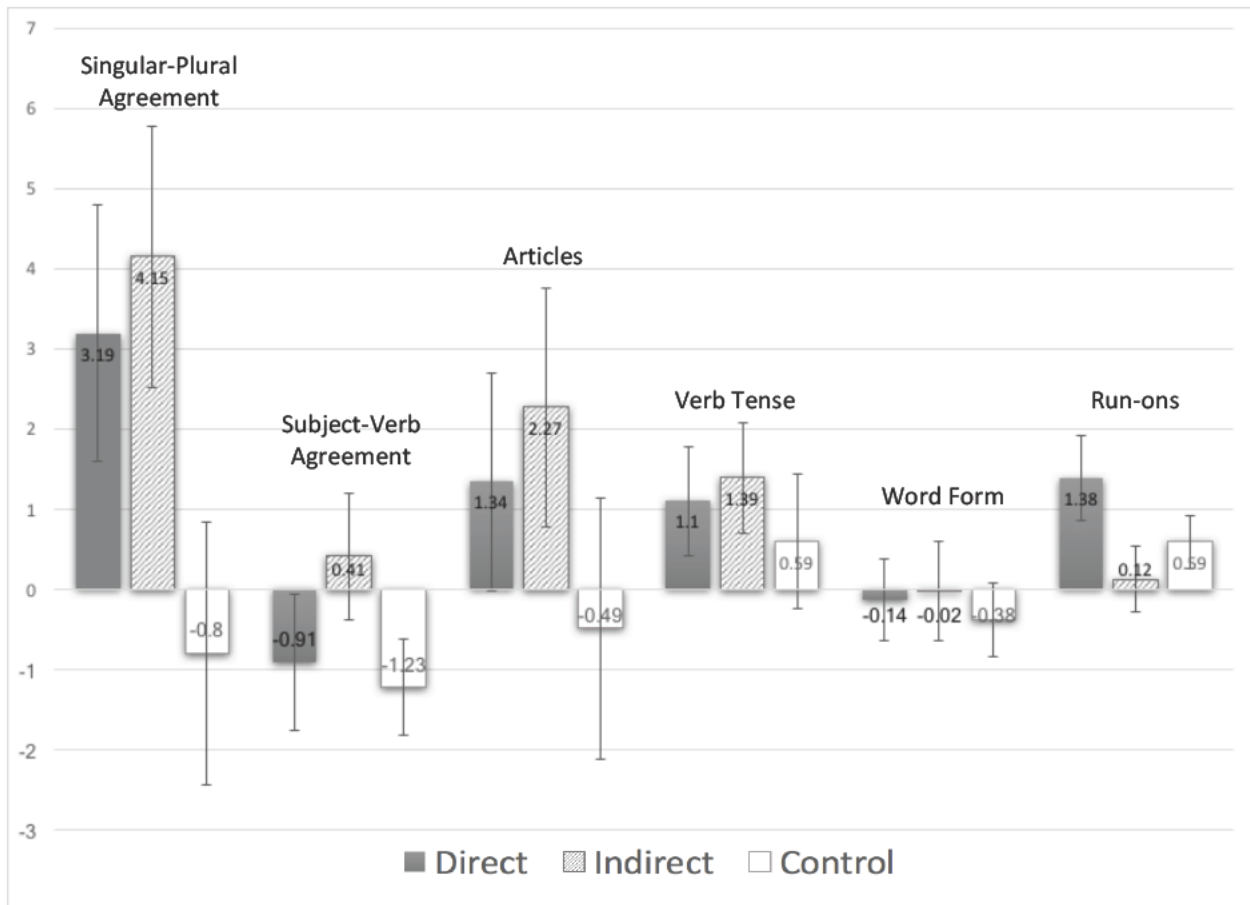


Figure 4. Group gains in accuracy by error type

were higher for indirect feedback ( $M = 4.14$ ,  $SD = 10.84$ , Cohen's  $d = .40$ ) than direct ( $M = 3.19$ ,  $SD = 10.22$ , Cohen's  $d = .28$ ), although not significantly from each other. Following Plonsky and Oswald's (2014) benchmarks for within-group design in L2 research, both treatment types had small effect sizes for gains in singular-plural noun agreement and were both significantly more effective than the control ( $M = -.80$ ,  $SD = 10.66$ , Cohen's  $d = -.08$ ) with high variation within each group. For each of the other error types, although a general trend appears of greater gains for the treatment groups than the control, only a few differences are significant. For example, for errors in subject-verb agreement, although both the gains for direct ( $M = -.91$ ,  $SD = 5.45$ , Cohen's  $d = -.22$ ) and indirect feedback ( $M = .41$ ,  $SD = 5.28$ , Cohen's  $d = .10$ ) were greater than the control ( $M = -1.23$ ,  $SD = 3.90$ , Cohen's  $d = -.29$ ), only the indirect treatment showed a

significant but small, as revealed by the non-overlapping error bars in Figure 4. Interestingly, both the direct and the control groups decreased in accuracy slightly for subject-verb agreement. For errors in subject-verb agreement, along with singular-plural noun agreement, articles, and word forms, the control group decreased in accuracy between the pre- and posttests.

The only other error type that resulted in significant differences between groups was run-on sentence errors, for which the direct group ( $M = 1.38$ ,  $SD = 3.39$ , Cohen's  $d = .41$ ) outperformed the indirect group ( $M = .12$ ,  $SD = 2.70$ , Cohen's  $d = .06$ ), but this difference was moderate and not statistically different from the control group ( $M = .59$ ,  $SD = 2.11$ , Cohen's  $d = .23$ ).

With respect to errors in articles, the indirect group ( $M = 2.27$ ,  $SD = 9.92$ , Cohen's  $d = .29$ ) experienced slightly greater gains than the direct group ( $M = 1.34$ ,  $SD = 8.72$ , Cohen's  $d = .16$ ) with small differences in effect sizes, while both groups experienced greater gains than the control ( $M = -.49$ ,  $SD = 10.57$ , Cohen's  $d = -.05$ ), although not statistically different. For errors in verb tense and aspect, the control group showed slight improvement without feedback ( $M = .59$ ,  $SD = 5.43$ , Cohen's  $d = .13$ ), suggesting that practice may have more of a positive effect for tense and aspect, while the indirect group ( $M = 1.39$ ,  $SD = 4.58$ , Cohen's  $d = .36$ ) performed similarly to the direct group ( $M = 1.10$ ,  $SD = 4.39$ , Cohen's  $d = .29$ ), both with small effect sizes. Finally, none of the groups improved in accuracy for word form (i.e., part of speech) errors, and the groups did not differ statistically, although mean scores for both the direct ( $M = -.14$ ,  $SD = 3.27$ , Cohen's  $d = -.05$ ) and indirect groups ( $M = -.02$ ,  $SD = 4.09$ , Cohen's  $d = -.01$ ) were higher than the control ( $M = -.38$ ,  $SD = 2.99$ , Cohen's  $d = -.12$ ), suggesting some minimal effect of feedback.



Overall, it appears that the error types that were most amenable to feedback in general in this data set were singular-plural noun agreement and articles (relative to the control group), and these were also the most common error types that occurred for this population. Singular-plural errors stood out in terms of effects for both feedback types, while run-on sentence and subject-verb agreement errors appeared to be more amendable to a particular feedback type (direct and indirect, respectively).

### **Results of Research Question 3: The Relationship between Grammatical knowledge and Effectiveness of WCF**

To investigate the potential interaction between students' grammatical knowledge and their utilization of feedback in general, gain scores in overall accuracy were correlated with overall grammatical knowledge test scores. The assumptions for Pearson's correlation were deemed sufficient by checking linearity and normal distribution through a visual inspection of scatter plots for each of the comparisons. In comparing gains with overall GKT results (across error types), no outliers appeared in the scatterplot for overall gains, however, an outlier appeared in each of the scatterplots for verb tense/aspect and run-on sentences. Likewise, comparing scatterplots of gains with metalinguistic knowledge test scores in isolation, an outlier was identified in the part of speech and the run-on sentence scatterplots. Each of these outliers was checked and deemed valid data points and were kept in the analysis. Both the gain scores and the test scores were converted to Z-scores to compare scores on different scales. Correlations were then analyzed between the overall gain Z-scores and the overall grammatical knowledge test Z-scores to determine if a relationship exists.

First, the descriptive statistics for results of the Grammatical Knowledge Test (including mean scores, standard deviation, skewness, and kurtosis) are presented by sub-section and for the overall test in Table 17. Two rounds of piloting the test in the target context allowed for improvements in the test design to reach a desirable overall difficulty of 60.5% to spread test takers' results for comparison. The Error Correction section (72.5%) was predictably the easiest, while the Metalinguistic Knowledge section (53.3%) was predictably the most difficult of the three sections. Next, the individual item analysis is presented in Table 18 with item difficulty, standard deviation, item discrimination, skewness, and kurtosis. Although the performance of a few of the items were less than desirable relative to the pilot results in terms of difficulty (e.g., item #4 for articles in the error correction section was answered correctly by all participants) and item discrimination (e.g. negative values resulted for item #14 in the error correction section and for item #1 in the error type recognition section), the overall performance was satisfactory. Reliability of the Grammatical Knowledge Test was high overall (Cronbach's alpha = .85), and satisfactory within each sub-section of the test, as presented in Table 19.

Reliability for the scoring procedures for the Metalinguistic Knowledge section of the test was checked by scoring each of the 12 Metalinguistic Knowledge items for 50% of the data set by a second rater (an undergraduate research assistant with coursework in linguistics). The inter-rater coding procedures began with an initial training session to discuss several examples of student responses for each item. Following this session, the second rater completed a first round of independent coding of 10% of the data set and was instructed to take notes on limitations of the rubric and student responses that did not readily match one of the three categories in the rubric. In a follow-up training session the raters discussed discrepancies in coding in the first 10% and collaboratively revised the scoring rubric to more reliably accommodate the range of

Table 17.

*Descriptive Statistics for the Grammatical Knowledge Test Results (by section)*

Test Section	Total Items	Total Points	Item Difficulty (% correct)	SD	Skewness	Kurtosis
Error Correction	24	24	72.45	14.61	-1.10	1.02
Metalinguistic Knowledge	12	24	53.32	20.81	0.38	-1.35
Error Type Recognition	36	24	55.64	15.38	0.47	0.05
Total	72	72	60.47	14.02	-0.01	-0.85

Table 18

*Grammatical Knowledge Test Item Analysis*

Subtest	Item Number	Error Type	Mean (Item Diff.)	SD	Item Discr.	Skewness	Kurtosis
Meta-linguistic Knowledge	1	VT (1/3)	.50	.84	.40	.00	-1.57
	2	SV (1/2)	.77	.81	.29	-1.28	-.23
	3	VT (2/3)	.46	.79	.33	.13	-1.36
	4	Art (1/3)	.67	.73	.26	-.62	-.87
	5	POS (1/2)	.57	.87	.44	-.30	-1.62
	6	SP (1/1)	.64	.72	.36	-.47	-.95
	7	RO (1/1)	.40	.89	.28	.38	-1.63
	8	Art (2/3)	.38	.65	.46	.27	-.71
	9	SV (2/2)	.48	.86	.30	.09	-1.66
	10	POS (2/2)	.51	.88	.43	-.04	-1.71
	11	VT (3/3)	.49	.64	.29	.01	-.53
	12	Art (3/3)	.44	.81	.43	.21	-1.43
	1	VT (1/6)	.55	.50	.20	-.21	-1.98
	2	SV (1/4)	.85	.36	.10	.15	-1.11
	3	VT (2/6)	.63	.49	.26	-.53	-1.75
	4	Art (1/6)	1.00	.00	n/a	n/a	n/a
	5	POS (1/4)	.70	.46	.36	-.89	-1.23
	6	SP (1/2)	.78	.41	.26	-1.38	-.09
	7	RO (1/2)	.55	.50	.28	-.21	-1.98
	8	Art (2/6)	.69	.46	.39	-.84	-1.31

Error Correction	9	SV (2/4)	.72	.45	.15	-1.00	-1.02
	10	POS (2/4)	.63	.48	.43	-.56	-1.71
	11	VT (3/6)	.84	.37	.25	-1.84	1.41
	12	Art (3/6)	.50	.50	.26	.01	-2.03
	13	SP (2/2)	.93	.25	.11	-3.47	10.16
	14	VT (4/6)	.63	.48	-.09	-.56	-1.71
	15	POS (3/4)	.71	.46	.28	-.92	-1.17
	16	SV (3/4)	.81	.40	.17	-1.58	.51
	17	VT (5/6)	.80	.40	.12	-1.48	.20
	18	RO (2/2)	.67	.47	.39	-.71	-1.51
	19	Art (4/6)	.62	.49	.06	-.50	-1.78
	20	SV (4/4)	.60	.49	.07	-.41	-1.86
	21	Art (5/6)	.80	.40	.27	-1.48	.20
	22	Art (6/6)	.80	.40	.19	-1.54	.37
23	VT (6/6)	.58	.50	.33	-.32	-1.92	
24	POS (4/4)	.60	.49	.19	-.41	-1.86	

Subtest	Item Number	Error Type	<i>M</i> (Item Diff.)	<i>SD</i>	Item Discrim.	Skewness	Kurtosis
Error Type Recognition	1	SP (1/3)	.52	.50	-.02	-.07	-2.02
	2	Art (1/9)	.61	.49	.32	-.47	-1.81
	3	VT (1/9)	.48	.50	.16	.07	-2.02
	4	VT (2/9)	.66	.48	.18	-.68	-1.56
	5	Distr (1/4)	.70	.46	.35	-.89	-1.23
	6	POS (1/6)	.56	.50	.28	-.26	-1.96
	7	SV (1/6)	.53	.50	.23	-.12	-2.01
	8	Art (2/9)	.67	.47	.29	-.75	-1.46
	9	Distr (2/4)	.52	.50	.25	-.07	-2.02
	10	Art (3/9)	.67	.47	.23	-.71	-1.51
	11	Art (4/9)	.54	.50	.32	-.18	-2.00
	12	VT (3/9)	.53	.50	.03	-.12	-2.01
	13	POS (2/6)	.64	.48	.32	-.59	-1.68
	14	Art (5/9)	.57	.50	.22	-.29	-1.94
	15	RO (1/3)	.44	.50	.23	.26	-1.96
	16	VT (4/9)	.65	.48	.13	-.65	-1.60
	17	POS (3/6)	.64	.48	.28	-.59	-1.68
	18	Art (6/9)	.50	.50	.28	-.01	-2.03
	19	VT (5/9)	.44	.50	.12	.26	-1.96
	20	SV (2/6)	.61	.49	.23	-.44	-1.84
	21	VT (6/9)	.76	.43	.06	-1.24	-.46
	22	POS (4/6)	.51	.50	.36	-.04	-2.03
	23	Distr (3/4)	.78	.41	.37	-1.38	-.09
	24	RO (2/3)	.42	.50	.09	.32	-1.92
	25	Art (7/9)	.42	.50	.22	.32	-1.92

26	SV (3/6)	.52	.50	.31	-.10	-2.02
27	Art (8/9)	.59	.49	.33	-.38	-1.88
28	VT (7/9)	.73	.44	.26	-1.07	-.86
29	POS (5/6)	.47	.50	.32	.12	-2.01
30	SP (2/3)	.66	.48	.26	-.68	-1.56
31	SV (4/6)	.56	.50	.28	-.24	-1.97
32	SV (5/6)	.50	.50	.34	.01	-2.03
33	Distr (4/4)	.87	.34	.26	-2.23	3.03
34	VT (8/9)	.59	.49	.15	-.35	-1.90
35	Art (9/9)	.24	.43	.08	1.20	-.57
36	RO (3/3)	.44	.50	.13	.24	-1.97
37	POS (6/6)	.67	.47	.14	-.75	-1.46
38	SP (3/3)	.51	.50	.11	-.04	-2.03
39	VT (9/9)	.61	.49	.18	-.44	-1.84
40	SV (6/6)	.65	.48	.38	-.65	-1.60

Table 19

*Reliability for Grammatical Knowledge Test (Cronbach's Alpha) by Test Section*

Subtest	Items	Reliability (Cronbach's alpha)	Cronbach's alpha (based on standardized items)
Metalinguistic Knowledge	12	.73	.74
Error Correction	24	.60	.60
Error Type Recognition	36	.75	.76
Total Test	72	.85	.84

responses in the sample (see Appendix A for the final version). After two training sessions and the revision of the rubric, an additional 40% of the data was coded independently by the second rater. Cohen's kappa (which incorporates agreement by chance in inter-rater agreement) was then run to determine the reliability of agreement in scoring using the customized scoring rubrics. There was strong agreement between the two raters,  $k = .94$ ,  $p = .00$ . Table 20 displays Cohen's kappa for inter-rater agreement in scoring for the metalinguistic knowledge section of

Table 20

*Inter-rater Reliability for Metalinguistic Knowledge Scoring*

Item Number	Error Type	Cohen's kappa
1	VT (1/3)	.98
2	SV (1/2)	.90
3	VT (2/3)	.95
4	Art (1/3)	1.00
5	POS (1/2)	.87
6	SP (1/1)	.90
7	RO (1/1)	.95
8	Art (2/3)	.92
9	SV (2/2)	.95
10	POS (2/2)	.89
11	VT (3/3)	.92
12	Art (3/3)	.98

the test by item for 40% of the dataset. Inter-rater agreement was high across all the error Categories with 95% significance of .00 for all 12 items. To test for equivalence in grammatical knowledge between the three groups, the overall test scores were subjected to a one-way ANOVA. Results showed that there was no significant difference between overall test results between the three groups,  $F(2, 124) = 1.71, p = .186, \eta^2 = .027$ .

Research question 3 sought to determine the role of grammatical knowledge (comprised of the sub-constructs of metalinguistic knowledge, error correction, and error type recognition) in moderating the overall effectiveness of WCF. Z-scores were generated for both overall Grammar Knowledge Test scores and for overall gains scores in accuracy. Pearson's correlation analyses were performed on the gain Z-scores and the overall Grammatical Knowledge Test scores, and then separately on the gain Z-scores and the Metalinguistic Knowledge scores. The results reveal that no significant correlations existed for all groups combined between overall gains in accuracy and overall grammar knowledge,  $r = -.09, p = .34$ , or between overall gain scores in accuracy and metalinguistic knowledge in isolation,  $r = -.12, p = .19$ . Likewise, comparing Pearson's correlation between each of the three groups, no significant relationships were found within groups,  $r = .08, p = .64$  (direct),  $r = -.27, p = .08$  (indirect),  $r = -.08, p = .60$  (control), although evidence of a small negative relationship is present for the indirect group, meaning that indirect feedback may limit students who come equipped with grammatical knowledge. Due to such limited associations found in the correlation coefficients for overall grammatical understanding of these structures and total gains in accuracy over the five weeks, a multiple regression model was deemed unnecessary as it would not provide further information about the interaction between variables in moderating gains in overall accuracy.

The next section presents results of correlations between gains in accuracy for each error type and the corresponding grammatical knowledge for each error type according to feedback type to determine a potential interaction between grammatical knowledge, feedback type, and the relative effectiveness of each feedback type.

### Results of Research Question 4: Interaction between Grammatical Knowledge, Feedback Type, and Development in Grammatical Accuracy

The final research question sought to determine the extent to which grammatical understanding of individual linguistic structures played a role in moderating the utilization of different types of WCF (direct vs indirect). It was hypothesized that when students are equipped with greater understanding of a specific grammatical structure, in terms of explicit and/or implicit understanding (Ellis, 2004), they would benefit more from receiving indirect feedback, while direct feedback would be more helpful for structures that are less understood, assuming less ability to successfully make use of the code (in conjunction with the grammar key and training).

First, results of the Grammatical Knowledge Test are presented according to individual error type, both for the overall test scores in Table 21, and for the metalinguistic knowledge sub-

Table 21

*Grammatical Knowledge Test Results (by error type)*

Error Type	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
SP	68.46	19.94	-.79	1.14
Art	58.47	17.80	-.37	-.26
SV	65.66	19.65	-.06	-.86
POS	59.41	23.30	-.22	-.79
VT	59.26	16.23	-.38	-.10
RO	51.57	32.05	-.03	-1.32
Total	60.47	14.02	-.26	-.78



section scores in isolation (Table 22), as Z-scores were computed in SPSS for each set of scores to compare with Z-gain scores in accuracy.

Table 22

*Metalinguistic Knowledge Section Test Results (by error type)*

Error Type	<i>M</i> (%)	<i>SD</i>	Skewness	Kurtosis
SP	63.78	37.10	-.49	-1.03
Art	48.68	25.68	.04	-.39
SV	62.60	29.36	-.20	-.75
POS	51.97	32.37	.01	-1.01
VT	49.61	26.10	.10	-.66
RO	43.31	44.70	.27	-1.71
Total	53.32	53.32	-.41	-.25

Finally, Pearson’s correlation analyses are presented on the individual gain Z-scores in grammatical accuracy for each individual error type and the corresponding Z-scores of grammatical knowledge for each error type. A matrix of correlation coefficients is presented in Table 23 to present the results. In line with the overall results, no clear patterns were found in relationships between individual error categories and grammatical knowledge for those grammatical structures. Oddly, the only significant correlation in the matrix in Table 23 showed that the more understanding students in the control group showed on the test for singular-plural noun agreement, the worse their accuracy became over the five weeks without feedback.

Results of the correlation matrices generally suggest that indirect feedback may be negatively correlated with overall grammatical knowledge ( $r = -.27, p = .08$ ), with four out of the six error categories showing a weak negative relationship (while errors in subject-verb agreement and run-on sentences showed a weak but statistically insignificant positive relationship). With a larger sample and greater statistical power (or perhaps a longer treatment period), it may be that a weak negative relationship could be observed between grammatical knowledge and gains in accuracy for the indirect WCF group. Considering that the indirect group experienced the highest mean gains of all groups, this could suggest that in this population of learners, students better

Table 23

*Correlation Matrix, Grammatical Knowledge Z-Scores x Gains in Accuracy Z-Scores (Pearson, 2-tailed)*

	Direct CF		Indirect CF		Control		Total	
	Correlation	Sig.	Correlation	Sig.	Correlation	Sig.	Correlation	Sig.
SP	.12	.46	-.17	.26	-.39*	.01	-.14	.11
Art	.02	.90	-.18	.23	-.19	.23	-.12	.17
SV	-.17	.30	.15	.35	.28	.08	.02	.79
POS	.19	.24	-.19	.23	.13	.40	.03	.71
VT	-.07	.68	-.12	.46	.08	.62	-.02	.79
RO	-.05	.77	.02	.90	.18	.26	.05	.57
Total	.08	.64	-.27	.08	-.08	.60	-.09	.34

equipped with grammatical knowledge may benefit relatively less from indirect feedback—contrary to the hypothesis. This remains speculation in light of the current results in this data set.

Pearson’s correlation analyses were also performed on the gain Z-scores for each error category and the corresponding Z-scores for metalinguistic knowledge (one of the three subsections of the Grammatical Knowledge Test). Metalinguistic knowledge was separated as a secondary analysis because of the theoretical arguments that have been made for explicit metalinguistic knowledge as potential moderator of the utilization of WCF (e.g., Ferris & Roberts, 2001; Park et al., 2016; Sheen, 2007). The correlation matrix for metalinguistic knowledge and gains in accuracy is presented in Table 24, which also reveals no significant pattern in the data. While no statistically significant relationships exist, it is noteworthy that

Table 24

*Correlations: Metalinguistic Knowledge Z-Scores x Gains in Accuracy Z-Scores*

	Direct CF		Indirect CF		Control		Total	
	Pearson Corr.	Sig. (2-tailed)	Pearson Corr.	Sig. (2-tailed)	Pearson Corr.	Sig. (2-tailed)	Pearson Corr.	Sig. (2-tailed)
SP	.22	.16	-.24	.11	-.30	.052	-.13	.15
Art	-.11	.50	-.10	.53	-.20	.21	-.14	.11
SV	.03	.87	.13	.41	.04	.79	.04	.66
POS	.27	.09	-.25	.10	.20	.20	.06	.51
VT	.04	.82	-.14	.37	.08	.62	.00	.99
RO	-.06	.72	.08	.59	.18	.26	.07	.46
Total	.10	.54	-.29	.06	-.13	.40	-.12	.19

correlation coefficients suggest evidence for opposite patterns with small relationships for a few of the error categories. For singular-plural noun agreement errors, indirect feedback resulted in a

weak positive relationship between gains and metalinguistic knowledge ( $r = .22, p = .16$ ), while direct feedback resulted in a weak negative relationship ( $r = -.24, p = .11$ ). The control group matches the pattern in the indirect group for singular-plural agreement, but the contrast with the direct feedback group could be worth investigating further to determine if explicit feedback is more helpful when learners are knowledgeable of this structure. A similar pattern is suggested for word form errors, which could point to either a potential advantage for direct feedback when grammatical knowledge is present for these features, or an advantage for indirect feedback when learners do not understand the grammatical rules for these features.

### **Post Hoc Analysis**

Due to the unusually high variation found at the group level throughout this data set (i.e., standard deviation values two to three times the size of mean gains), questions remain concerning a potential interaction between grammatical knowledge, linguistic structure, and the explicitness of feedback type in the development of accuracy. Such high variation in gains within groups likely points to the complex relationship between corrective feedback and development in accuracy in second language writing. As illustrated in the literature review, several factors likely influence the effectiveness of WCF, and therefore isolating a single variable (grammatical knowledge in this case) presents challenges in research design. In light of such variation, a post hoc analysis of the data was carried out by separating students within each group according to their overall grammatical knowledge for each error category. The rationale is that with so many potential variables at play, focusing on those learners who more clearly represent the presence or absence of grammatical knowledge (i.e., grouping the extreme cases of the highest and lowest achievers on the test by each structure for comparison), may allow patterns to emerge in the variable in question.

Therefore, a post hoc analysis was conducted by focusing on the top and bottom quartile ( $n = 10$ ) of achievement on the grammatical knowledge test, both for overall grammatical knowledge and by individual structure. So, for example, the 10 students in each group who scored the highest overall on singular-plural noun agreement on the test were grouped together, and likewise for the lowest 10 scorers for each group. With groups of such limited size, statistical tests were not carried out in favor of presentation of descriptive statistics (means and standard deviations), which are presented in Table 25. Results show slightly lower variation relative to the full group level, and perhaps the most interesting finding from this perspective is

Table 25

*Post-hoc Analysis: Gains in Accuracy for Highest/Lowest Grammatical Knowledge Groups*

	Gram. Knowl.		S/P	S/V	Art	VT	WF	RO	Overall
Dir.	Highest	<i>M</i>	2.85	-1.15	5.43	1.46	0.58	0.50	10.26
	$n = 10$	<i>SD</i>	5.55	4.07	3.97	3.44	2.49	1.52	15.08
	Lowest	<i>M</i>	6.34	0.42	0.79	2.06	-0.22	0.65	10.01
	$n = 10$	<i>SD</i>	14.41	5.93	10.18	4.57	2.21	2.41	13.69
Ind.	Highest	<i>M</i>	-.14	1.04	-0.22	0.47	0.15	0.01	2.19
	$n = 10$	<i>SD</i>	9.11	6.49	11.48	2.47	2.77	1.56	16.09
	Lowest	<i>M</i>	6.45	0.36	3.36	3.50	1.24	0.18	11.07
	$n = 10$	<i>SD</i>	8.94	5.76	11.48	5.72	3.43	3.34	16.52
Con.	Highest	<i>M</i>	-5.58	-0.82	-4.30	2.81	0.33	1.68	-7.40
	$n = 10$	<i>SD</i>	13.70	3.61	12.04	6.21	2.80	2.28	20.75
	Lowest	<i>M</i>	3.20	-2.69	5.34	0.53	-0.02	0.49	-.05
	$n = 10$	<i>SD</i>	13.70	4.88	11.44	5.87	2.68	1.92	21.25

that although the students with lowest grammatical knowledge in the indirect and control groups experienced the most gains (which makes intuitive sense, as less understanding suggests more

room for development in accuracy), those in the direct group with high grammatical knowledge experienced similar overall gains as those with the least knowledge. Figures 5 (overall) and 6 (for articles) present visual representations of the changes in accuracy in writing between the pre- and posttest for individual students in these extreme groups. Note that the data points in these figures represent error rates at the pre- and posttests, therefore a downward slope represents improvement in accuracy (decrease in error rate). Figure 5 shows that the contrast in performance between students with the highest overall grammatical knowledge depending on feedback type. Those with the highest overall grammatical knowledge (in the top quartile) barely decreased in their error rates with a mean gain of 2.19 and high variation ( $SD = 16.09$ ), as depicted in the figure on the left. Meanwhile, the top quartile performers in overall grammatical knowledge in the direct feedback group experienced mean gains in accuracy of 10.26 ( $SD = 15.08$ ), demonstrated by the downward slopes in the figure graph on the right in Figure 5.

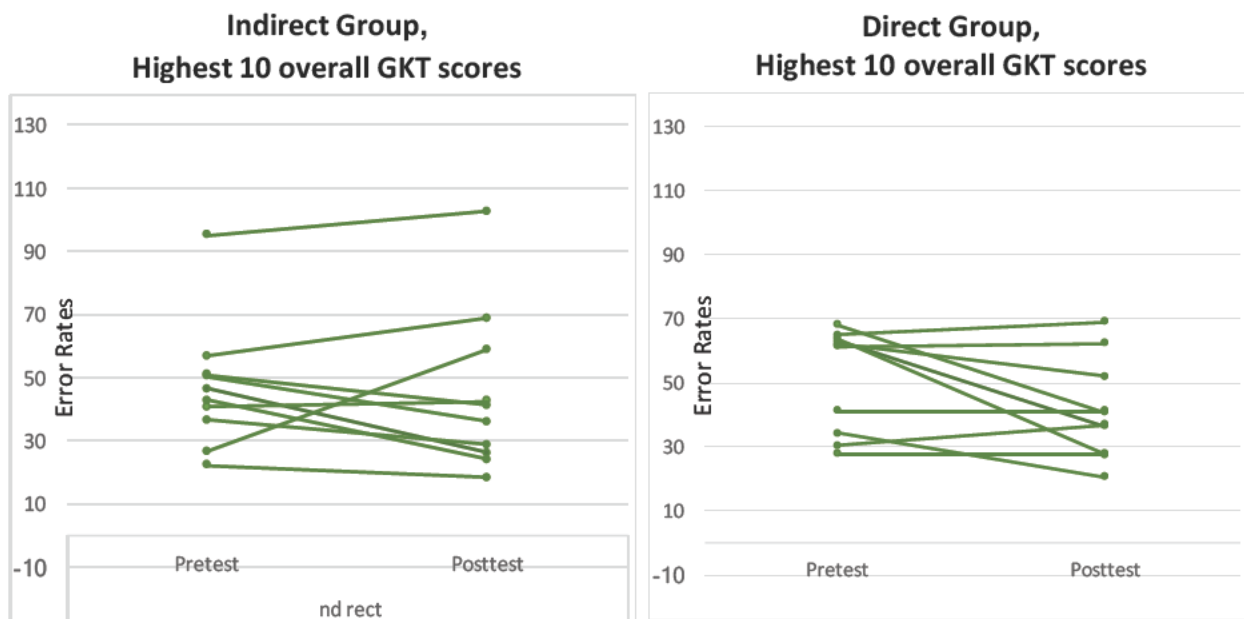
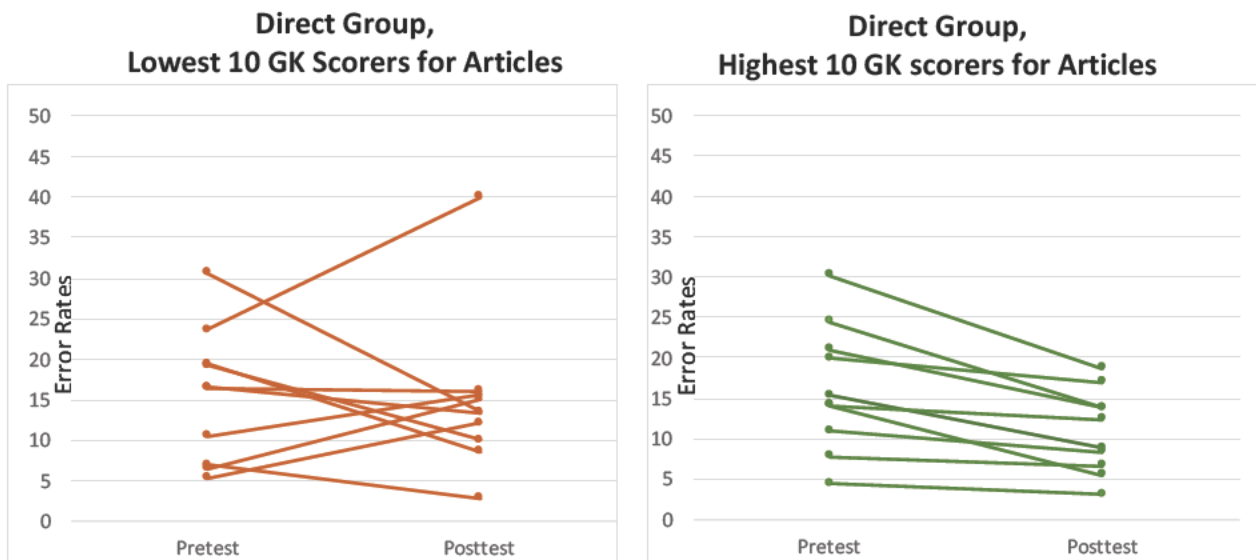


Figure 5. Gains in overall accuracy comparing highest overall test performers between feedback types.

An examination of individual error categories shows this pattern most prominently for articles, where the highest scorers for knowledge of articles appeared to outperform the lowest scorers ( $M = 5.43$ ,  $SD = 3.97$  and  $M = .79$ ,  $SD = 10.18$ , respectively). This pattern can be seen in Figure 6 in which the students with the lowest grammatical knowledge scores for articles (on the left) showed wider variation, while the those with the highest knowledge of articles (on the right) appear to be improving in a more consistent pattern. Although comparisons of groups of 10 offer limited confidence, these patterns are promising and worth investigating with further research. If the patterns hold with larger groups of high and low grammatical knowledge, it could suggest



*Figure 6.* Gains in accuracy for articles comparing lowest and highest test performers in the direct feedback group

that for students who are better equipped with grammatical knowledge, direct feedback may be more helpful than indirect, particularly for articles. As will be discussed in the next section, the difficulty or structural complexity of error types could play a role if certain error types can benefit in accuracy development despite students' substantial knowledge of the structures before receiving feedback.

## Summary of Results

To summarize the findings in this study, in response to the research question 1, WCF was found to be effective (for either treatment type) relative to a control group in terms of developing grammatical accuracy in new writings over time. Both the direct and indirect feedback types statistically outperformed the control group in developing grammatical accuracy with small to moderate effects, although the groups were not statistically different from each other in terms of gains.

In response to research question 2, the linguistic target appears to play a role in the moderating of WCF. Both feedback types appeared to benefit the development in accuracy for errors in singular-plural noun agreement the most, as both treatment groups gained significantly more in accuracy compared to the control group. Findings for each error type also suggest that indirect feedback could be more beneficial than direct for subject-verb agreement errors, while errors in run-on sentences (the only syntactic category) favored direct feedback over indirect.

Considering the potential role of grammatical knowledge in moderating the effectiveness of WCF (research question 3), results at the group level showed no influence of grammatical knowledge on gains in accuracy over the course of this study with insignificant correlation both for overall grammatical knowledge scores and for metalinguistic knowledge scores.

In response to research question 4, no interaction was apparent at the group level between grammatical knowledge, feedback type, and gain scores across individual error types, as shown by correlation matrices. However, the variation in the data was quite high, suggesting that group level analysis may not be appropriate to isolate the variable of grammatical knowledge. A post hoc analysis of the top and bottom quartile of performers on the GKT for each error type



revealed a promising pattern that suggests students with the highest grammatical knowledge may benefit more from direct WCF, particularly in developing their accuracy in articles.

## **Chapter 5**

### **Discussion**

To advance our understanding of how WCF can be beneficial for L2 learners, researchers continue to explore the contextual variables that may moderate its effectiveness. This study investigated grammatical knowledge, which has been hypothesized as a potential moderating variable in the utilization of WCF. A test was developed to discover the potential role of grammatical knowledge in moderating learners' ability to make use of different types of corrective feedback to develop accuracy in new writings over time. In this study, 127 Thai EFL learners of English at an upper-intermediate writing proficiency were divided into three groups: one that received direct feedback, another that received indirect coded feedback, and a third that did not receive grammatical feedback on accuracy. The participants took a grammatical knowledge test and then wrote eight timed essays in class over a seven-week period, receiving feedback on five of the essays with the first two and last two serving as the pre- and posttests to measure rates of accuracy. In this chapter, the results that were outlined in Chapter 4 are interpreted in the context of previous findings to explain the issues under investigation in this study: (a) the effects of direct vs. indirect WCF relative to a control group, (b) the role of error type in moderating the effectiveness of WCF, relating to previous theoretical arguments that have categorized linguistic targets based on key features of learnability and complexity, (c) and the potential role of grammatical knowledge in moderating the use of WCF.

#### **Research question 1: Relative Effectiveness of WCF Types**

The findings in this study in response to the first research question provide further support for the efficacy of WCF in helping learners develop their written grammatical accuracy in new writings over the course of a semester. In this study, both direct and indirect feedback

types resulted in small to moderate effects after five treatments over five weeks. These findings are generally congruent with results in Kang and Han's (2015) meta-analysis of the effectiveness of WCF, particularly in view of the context of this study. Based on Kang and Han's (2015) sample of 22 studies that investigated the effectiveness of WCF in development over time, they found an overall moderate effect in developing grammatical accuracy (Hedges's  $g = .68$ ,  $SE = .13$ ,  $CI [.42, .99]$ ) with about 75% of the observed heterogeneity accountable to differences between studies ( $Q = 87.18$ ,  $p < .0001$ ). Although the current study lacked a delayed posttest, previous studies that have used a delayed posttest indicate moderate effects for learning retention (Hedges's  $g = .68$ ,  $SE = .095$ ,  $CI [.50, .87]$ ,  $p < .0001$ ). Kang and Han analyzed study variables as moderators, and although the sample size was limited, they found less effectiveness for WCF in foreign language settings ( $k = 6$ ,  $g = .22$ ,  $CI [.00, .44]$ ) than in second language settings ( $k = 15$ ,  $g = .66$ ,  $CI [.44, .89]$ ), which suggests relatively greater effect of WCF in this study compared to other FL contexts. They also found less effectiveness for unfocused (i.e., covering a range of error types) WCF ( $k = 9$ ,  $g = .33$ ,  $CI [.14, .52]$ ) than for focused WCF ( $k = 10$ ,  $g = .69$ ,  $CI [.35, 1.03]$ ), and oddly, less effectiveness the more treatment sessions were provided (for one shot design,  $k = 5$ ,  $g = .83$ ,  $CI [.32, 1.34]$ , and for more than three treatments,  $k = 10$ ,  $g = .31$ ,  $CI [.13, .50]$ ), although this could be explained by the fact that most of the studies that focused on only one error type also employed a single-shot design.

In comparing direct and indirect feedback types, findings in this study are also comparable to Kang and Han's (2015) findings in that both types were effective but not statistically different from each other. They reported moderate effect for direct feedback ( $k = 12$ ,  $g = .60$ ,  $CI [.42, .77]$ ) and small effect for indirect ( $k = 5$ ,  $g = .36$ ,  $CI [.04, .69]$ ), although again it should be noted that many of the studies that focused on direct feedback also focused on a single

error type, which has been found to inflate effect sizes (e.g., see the series of studies by Bitchener and colleagues). Although both the results of this study and the aggregate results in Kang and Han (2015) reveal no significant difference between direct and indirect WCF types, differences in the indices of practical significance are notable. The results in this study suggest benefits for the indirect approach with greater effect sizes for most of the error types, although after five treatments the effects were small relative to findings in L2 research in general, particularly for studies using within-group designs (Plonsky & Oswald, 2014). Van Beuningen et al. (2012) also found a slight advantage for indirect feedback over direct for gains that persisted in delayed posttests. Their study was on Dutch FL students at the high school level using a comprehensive approach.

If indirect WCF were to have true advantage for the sample in the present study (i.e., with greater power), study variables could help explain differences relative to the trends in favor of direct WCF in Kang and Han's (2015) meta-analytic findings. For instance, the Thai students in this study were from elite educational institutions in Thailand and their general understanding of grammar and motivation to improve their written accuracy were quite high. Another factor that could have contributed to a potential advantage for the indirect feedback in this study was that the scope of the feedback was relatively limited (six categories as opposed to a wider range of symbols/categories in previous "unfocused" designs). It may be that the use of an indirect code is easier to internalize and handle when the number of categories is limited, rather than covering a wide range of error types (e.g., over a dozen different categories). Additionally, previous studies have frequently failed to report whether students in an indirect group are provided with an error key (44% unreported) to help them make sense of coded feedback they receive, according to Liu and Brown's (2015) methodological synthesis of 44 studies in this area. Likewise, only 16% of

the studies in their sample reported training to help students make use of an indirect code, which may also explain relatively greater gains for indirect WCF in this study compared to the aggregate results of Kang and Han (2015). Other study variables may also help explain this difference, such as students' past experiences with particular feedback strategies or their individual preferences. A larger pool of studies in future meta-analyses that include indirect WCF groups in different educational contexts could shed light on moderator analysis and help explain the factors that affect the relative usefulness of different feedback types.

Overall, the findings in this study generally support those of recent studies with improved methodological design (e.g., Kang & Han, 2015; van Beuningen, 2012) providing further evidence that feedback helps in the development of accuracy, resulting in small gains after a handful of treatments. The current study benefited from an overview in methodological design in WCF research (Liu & Brown, 2015) employing a relatively large sample size and larger quantity of writing under analysis (over 800 word samples for each of the pre- and posttests) compared to most other studies (typically with pre- and posttest samples between 100-300 words). This study also controlled for proficiency, instruction in grammar, training to make use of the feedback, exposure to a code for the indirect group, time restriction, and the use of reference tools during writing. Although this study lacked a delayed posttest, the posttest was comprised of two writings that occurred about a week apart, and separate analysis of these two essays revealed no statistical difference. These results, along with those of other recent WCF studies (as design has improved), should help end the debate over the general effectiveness of WCF that Truscott proposed in 1996 when he argued that WCF is not only ineffective in promoting the development of accuracy, but is counterproductive for L2 learners. On the other hand, taking into consideration the means required to achieve these effects (e.g., the investment of time and effort

for teachers), the benefit of WCF may be put into question. In the context of this ongoing debate, it could be argued that even a small effect after just five treatments represents substantial development (if the effects are indeed persistent over time). Despite the great effort that WCF involves, writing teachers can be somewhat confident to continue to provide WCF to students who aim to improve their written accuracy, although practitioners, policy-makers and students would be wise to limit expectations of the effectiveness of WCF within a single semester or after just a few writing assignments.

Regarding the potential advantages for feedback type, no clear distinctions could be made in this study that point to the overall benefit of more or less explicit feedback types (if practitioners are hoping to choose one or the other) as both were helpful. It may be that other variables, such as student or teacher preference could be the most important factor in deciding the most useful feedback type, particularly if a single approach is taken in providing WCF.

It was also noteworthy in these results that the control group decreased in overall accuracy between the pre- and posttest since previous research has shown gains in accuracy as a result of writing practice without feedback (e.g., Robb, Ross, & Shortreed, 1986; van Beuningen, et al., 2012). This decrease could have been a result of a more difficult posttest because of task type. Writing seven (the first half of the posttest) asked students to argue for or against the use of GMOs and/or pesticides by farmers in Thailand. This topic was likely less accessible for students than the other topics that made up the pre- and posttests which included part-time jobs, the use of mobile phone, and the use of laptops in school—all topics for which students are likely to have personal experience. As shown in Table 14, accuracy was slightly lower for writing seven on GMOs relative to the other posttest (writing eight on laptops). A superficial review of the writing data suggests that potential differences in patterns of noun use

between the writings (specifically, the ratio of regular to irregular nouns) could possibly explain the decrease in accuracy for the control group. Further lexical analysis using corpus tools could help provide insight and potentially help control for textual differences in future WCF research. For example, a fine-grained analysis of performance within typical error categories (as followed for the test development in this study) could reveal irregular features that behave differently in terms of accuracy, and these features could then be accounted for in analysis to better control for comparison of accuracy between topics.

### **Research Question 2: Interaction between WCF Type and Error Type**

In line with the limited research that focuses on linguistic structures, results of this study suggest an influence of error type on the development of accuracy in response to WCF, and some error types behaved differently in response to different WCF types. Descriptive and inferential statistics in this study suggest that at least one of the two feedback methods outperformed the control group for each error type, except part of speech errors. Statistically significant advantages for feedback were found for both WCF types on singular-plural agreement, for indirect WCF on subject-verb agreement, and for direct WCF on run-on sentences.

Truscott (1996, 2007) proposed that grammatical errors were impervious to WCF. Van Beuningen et al. (2012) is the only study to operationalize error type along these lines comparing grammatical and nongrammatical error types in the development of accuracy response to direct and indirect WCF. Results of the present study put into question van Beuningen et al.'s (2012) findings regarding the most suitable feedback type for grammatical errors. They used a one-shot treatment design with comprehensive feedback and found that only direct feedback led to development in accuracy for grammatical error types, which was also durable (medium magnitude) in a delayed posttest. The present study focused only on grammatical errors, but

found that indirect feedback had an equal or stronger effect as direct for five of the six targeted error types (all but run-on sentences). As mentioned earlier, study features could have led students to make better use of indirect feedback, such as training to use the code, consistent exposure to the code as it was pasted to all writings, or generally high grammatical knowledge for these students. These results suggest we should be cautious in concluding on advantages for direct feedback on grammatical error types. The present study design, however, lacks a delayed posttest, so it would be useful to know if the small to medium effects found for these students after five treatments would persist several weeks after receiving feedback, since a notable finding for van Beuningen et al. (2012) was that the direct (but not indirect) feedback persisted for the grammatical errors in the posttest. Although both feedback types were effective in this study, investigating differences in durability could help in guiding teachers' decisions, and was a limitation in this design.

Clause-based error types (subject-verb agreement and run-on sentences in this study), also classified as morphological and syntactic domains of linguistic knowledge in this study, have been found less susceptible to successful revision in response to feedback (e.g., Ferris, 2006; Ferris et al., 2000; Ferris & Roberts, 2001). Ferris and Roberts (2001) found subject-verb agreement errors as the least successfully revised error type in response to indirect feedback, although structurally rule-based and straightforward, and attributed this possibly to low communicative value (Ferris, 2006). Liu (2016) also found that clausal errors were consistently more often ignored by ESL students in revising, and suggested that development of accuracy in clausal errors may be more challenging as it requires deeper grammatical analysis compared with correcting word and phrase level errors. In this study, feedback type played a significant role for each of the clausal errors, although with opposite advantages.



Although direct and indirect feedback strategies were not statistically different from each other in developing accuracy in subject-verb agreement, indirect feedback was uniquely (moderately) more effective than no feedback (indirect: Cohen's  $d = .10$ ,  $CI [-1.19, 2.01]$ , control: Cohen's  $d = -.29$ ,  $CI [-2.44, -.01]$ ), whereas the effect of direct feedback did not differ significantly from either the indirect or control group (control: Cohen's  $d = -.22$ ,  $CI [-2.63, .81]$ ). This finding supports Ferris' treatability hypothesis, as the rule for subject-verb agreement is straightforward and likely learned early in this context. On the other hand, the finding that direct feedback was significantly more effective in the development of accuracy for run-on sentences (the other clausal feature in this study) than the indirect feedback was noteworthy as the only structure that appears to be more amendable to direct feedback, although the effect was not significantly different from the control (direct: Cohen's  $d = .41$ ,  $CI [.31, 2.45]$ , indirect: Cohen's  $d = .06$ ,  $CI [-.70, .94]$ , control: Cohens'  $d = .23$ ,  $CI [-.07, 1.25]$ ). Ferris (2006) found run-on sentence errors to be the only errors to regress in accuracy for ESL learners' revision in response to indirect feedback, so considering the current finding, it may be that learners require more explicit information that direct feedback provides to develop in this area. While these are considered a "treatable" error category by Ferris, if direct feedback is relatively more useful than indirect for run-ons, it contradicts Ferris et al. (2000) and other arguments (e.g., Park et al., 2016) that indirect feedback is beneficial for rule-governed structures. Although the correction required for most run-on sentence errors is straightforward (e.g., adding punctuation or a connector), it could be that learners rely less on rule-governed explicit knowledge of syntax to correct these types of errors, and more on a feel for what sounds correct once an error is pointed out (similar to the idiosyncratic correction of word choice or preposition errors). Although both clausal error types in this study (subject-verb agreement and run-on sentences) are "treatable" they could

behave differently due to structural complexity since run-ons relate to two clauses and subject-verb agreement does not.

The benefits (although moderate) of both types of feedback for singular-plural agreement was encouraging in this context (direct: Cohen's  $d = .28$ ,  $CI [-.03, 6.42]$ , indirect: Cohen's  $d = .40$ ,  $CI [.86, 7.45]$ , control: Cohen's  $d = -.08$ ,  $CI [-4.12, 2.52]$ ) as this represented the most frequent error type in the sample. These results contradicted Ferris et al.'s (2000) study, which found word form errors (mainly concerning singular-plural agreement) to progress only slightly. They attributed this to low communicative value, which could possibly explain differences in amenability to WCF in ESL vs. EFL contexts, particularly as accuracy was so important in the context of this study. Considering the rules for singular-plural agreement errors are usually straightforward, these errors may be more likely to be mistakes (Ferris, 2011). It could be useful for future research to perform a deeper analysis of error types, as irregular forms of singular-plural agreement could potentially respond differently to different feedback types. This will be discussed further as recommendations for future research.

Finally, at the lexical level, this study investigated articles and part of speech errors and found no statistically significant benefit for either feedback type compared to the control, although mean gains were higher for each of the treatment groups for these error types. Previous studies have found lexical errors to be "untreatable" (e.g., Bitchener et al., 2005; Ferris & Roberts, 2001), although evidence has been limited to revision. In terms of development, article errors have been the most widely studied linguistic target and have been found highly amenable to WCF with benefits for direct feedback in the development of accuracy, particularly in a series of controlled experiments in the last two decades (e.g., the series of studies from Bitchener and colleagues). Many of the studies, however, have focused only on article use, and as van

Beuningen et al. (2012) and Ferris (2011) argue, such clear focus likely leads to monitoring of the use of that single target feature in posttest performance, and is therefore not authentic to the needs to L2 learners in writing classrooms. Although no significant gains were found for articles in this study using a less focused approach, indirect feedback seemed helpful with moderate effect (Cohen's  $d = .29$ ,  $CI [-.74, 5.29]$ ) relative to the control (Cohen's  $d = -.05$ ,  $CI [-3.78, 2.81]$ ).

The few studies that examine linguistic target reveal that the structure plays a role in determining the effectiveness of WCF, but not enough is generalizable to guide pedagogical decisions. A clearer picture could potentially save time and effort of teachers and L2 writers by concentrating effort where it is most effectively used, while maintaining the goal of helping students improve overall accuracy. To better understand the amenability of different linguistic structures to WCF more studies are needed that isolate structures in data reporting (that include standard deviations or effect sizes for individual error types (Liu & Brown, 2015), particularly in studies that investigate development over time. Particularly because variation is so high in WCF research, future meta-analytic efforts will be needed to understand the multitude of variables that are likely at play. We currently have little understanding about what it is about particular linguistic targets that might cause some benefit from deeper processing and self-repair. Rather than taking an absolute position of treatability based on grammaticality of linguistic targets, it could help to advance this line of research by operationalizing taxonomies of learning difficulty for linguistic targets, such as Roehr and Gánem-Gutiérrez (2009). Their taxonomy considers schematicity (the extent to which a rule applies to all cases), conceptual complexity (how straightforward a grammatical rule is and how many elements are involved), technicality of metalanguage involved in a grammatical rule, and truth value (the extent to which a rule applies

in all cases). The consideration of these nuances in the amenability to different feedback types, as well as the role of other learner and contextual variables can shed more light on the role of error type.

### **Research Questions 3 and 4: The Relationship between Grammatical Knowledge and Effectiveness of WCF**

A central goal of this study was to determine the potential role of prior grammatical knowledge on (a) the development of accuracy in response to feedback, and (b) the relative amenability to different feedback types. Among several potential factors that could influence learners' use of WCF, grammatical knowledge (or metalinguistic knowledge) has been mentioned repeatedly in over decades of literature, though it has never been the focus of research design (Ferris & Roberts, 2001, included this variable but only in the context of revision). If the linguistic target does indeed influence the effectiveness of WCF, grammatical knowledge seems a likely moderator in learners' interaction with feedback they receive. And if teachers intuitively provide direct feedback for certain structures (Ferris, 2006) it could be useful to consider what their students know in judging their ability to self-repair in response to an error code. Anecdotal experience as an L2 teacher and teacher trainer suggests that experienced teachers often provide a mixed approach, likely extending indirect feedback when they sense students are equipped with the tools to make repairs themselves. The extent to which these decisions relate to linguistic structure or learner internal variables is an empirical question that this study attempts to address. There is also a clear need to study the effects of providing a mixed approach of WCF (relative to staying consistent between feedback types), which would lend ecological validity to this domain.

The only study to attempt to operationalize grammatical knowledge in the role of WCF was Ferris and Roberts (2001), which focused on revision in the same writing tasks. They found that grammatical knowledge correlated with ability to revise in response to indirect feedback for verb errors, noun ending errors, and wrong word errors, but not for article and sentence structure errors. This suggested that formal grammar knowledge appeared to play a role in the editing stage for all but perhaps the most complex structures.

The present study focused on the relationship between grammatical knowledge and development in accuracy (rather than revision). The central finding is that grammatical knowledge was not found to have had a significant relationship with the overall development of accuracy over the course of the semester for these learners. Pearson's correlation analyses were used initially to explore relationships between gains in accuracy and grammatical knowledge to determine if further analysis of regression was necessary. Correlation analysis revealed nearly no association between gains in accuracy and grammatical knowledge (or between gains in accuracy and metalinguistic knowledge), and while correlation analysis does not fully address the research questions related to the influence of grammatical knowledge on gains (or their interaction), a lack of association at this level signifies no need for an analysis of regression.

Because the variation in the gain scores in accuracy was quite large in all groups (standard deviation measures two or three times the mean gains), measuring the effects of grammatical knowledge at the group level might prove insufficient given the possibility of confounding variables at play (i.e., unmeasured learner and contextual variables that could mediate language development and/or the utilization of WCF). In other words, such variation makes for a cloudy picture when aiming to observe the effects of one of myriad potential variables at play. For this

reason, post hoc analysis was conducted to attempt to explore the variable of grammatical knowledge in isolation.

When isolating for this group characteristic in each of the groups, results point to promising patterns that suggest an influence of grammatical knowledge in these learners use of different feedback types. Descriptive statistics that compare the learners most and least equipped with prior grammatical knowledge ( $n = 10$  per group), reveal predictably that students who scored lowest on the grammar test experienced greater gains in accuracy than the highest scorers between the pre- and posttests. However, a promising pattern suggests that the highest scorers benefited more overall from direct feedback (overall gains of  $M = 10.26$ ,  $SD = 15.08$ ) than indirect ( $M = 2.19$ ,  $SD = 16.09$ ). This difference was most pronounced for articles, as roughly the top quartile of scorers on the test for article knowledge made greater gains in accuracy for articles in response to direct feedback ( $M = 5.43$ ,  $SD = 3.97$ ) and with less variation than the lowest scorers ( $M = .70$ ,  $SD = 10.18$ ). Gains in article accuracy for the highest scorers in the indirect group ( $M = -.22$ ,  $SD = 11.48$ ) seemed better than the control ( $M = -4.30$ ,  $SD = 12.04$ ) but not nearly to the extent of the direct feedback, and for the lowest scorers, effects of the indirect feedback ( $M = 3.36$ ,  $SD = 11.48$ ) were more consistent with those in the control ( $M = 5.34$ ,  $SD = 11.44$ ). English language teachers and learners are often aware of the difficulty in mastering accuracy in article use, particularly in spoken or timed writing contexts. According to Roehr and Gánem-Gutiérrez's (2009) taxonomy of learning difficulty described above, articles represent one of the most conceptually demanding structures in English (also low in schematicity and high in technicality of metalanguage involved in following the range of rules). Idiosyncratic elements involved in English article usage add to their complexity. If the pattern in this post hoc analysis were to hold true with larger groups—that direct WCF can be of greater help to learners

equipped with more grammatical knowledge on some error types—then learners who bring with them grammatical knowledge or awareness likely reach a threshold in terms of their ability to self-correct given their understanding of a structure. And especially for more complex structures (i.e., higher learning difficulty) for which errors are less likely to manifest as mistakes (as opposed to errors), direct feedback may be beneficial in guiding learners to benefit from the feedback they receive. It is worth noting that a large proportion of the research conducted on the efficacy of WCF in developing accuracy has focused exclusively on articles, which have shown the benefit of feedback, often direct. However, if articles prove to operate differently in response to feedback with unusual advantages for direct feedback for learners with more robust grammatical understanding, it is worth paying more attention to this potential moderating variable moving forward. Although no conclusions can be made about an interaction between metalinguistic knowledge, feedback type, and gains in accuracy at the group level from these findings, these trends motivate the need for further research. Grouping learners based on grammatical knowledge with larger sample sizes could be fruitful in better understanding the potential role of grammatical knowledge in the efficacy of WCF.

If grammatical knowledge does prove influential in providing the most effective WCF, how might this benefit second language teachers from a practical perspective? It could certainly be impractical in many teaching contexts to both measure learners' grammatical knowledge and then act on the results to tailor feedback techniques for individual students, especially in large classes with diverse students. However, if the pattern discovered in the post hoc analysis proves generalizable, we may find that a quick diagnostic of certain features (e.g., structurally complex but rule-governed) could realistically and fruitfully help teachers to tailor WCF for learners' readiness in many teaching contexts. For example, in smaller classes or individual writing

conferences, a tailored approach could be feasible and practical. Furthermore, as computer-mediated WCF is likely to come of age soon given the current pace of development in artificial intelligence for language processing, this type of aptitude-treatment interaction research will become crucial in designing automated feedback for individualized learning.

Many teachers opt for a mixed approach in providing their students WCF on grammatical accuracy following intuitive judgments on a case-by-case basis. Ferris (2006) discovered ESL teachers to provide direct feedback on the more idiosyncratic “untreatable” categories (even when they were asked to provide only indirect). If teachers likewise intuitively provide indirect feedback for “treatable” error types (which include articles), it may prove limiting if we can better understand the knowledge that students bring to a writing classroom. Findings in this area could help guide teacher strategies by informing our intuitive judgments as to which students, structures, and feedback type match as L2 writing teachers continue to spend countless hours providing WCF in hope of developing students’ accuracy in writing.



## **Chapter 6**

### **Conclusion**

#### **Theoretical Implications**

From a theoretical perspective, results of this study further support the role of negative evidence in L2 development in the context of writing. The finding that indirect feedback (which represents metalinguistic input with the use of a code) was at least as helpful as direct feedback in developing grammatical accuracy suggests that cognitive processes are at play, lending support for the role of explicit grammar knowledge in SLA for adult learners. Additionally, if the pattern observed in the post hoc analysis (i.e., that learners with higher grammatical knowledge could benefit more from direct feedback on certain structures) holds true with further investigation, it would show support for the role of explicit knowledge in the acquisition of implicit, proceduralized knowledge (Ellis, 2004).

Other theoretical arguments that have surfaced in the WCF literature, first from Truscott (1996, 2007) and later from Ferris (1999, 2004) and Park et al. (2016), posit the nature of linguistic structures and their role in susceptibility to feedback. Truscott's claim that only "nongrammatical" error types are amenable to development in response to WCF seems unfounded in light of these results and those of van Beuningen et al. (2012). Although this study did not compare grammatical and nongrammatical error types, several grammatical features were proven amendable to both direct and indirect feedback in this sample.

#### **Pedagogical Implications**

A bulk of the studies in this domain of research that have provided evidence in support of WCF in developing accuracy over time have focused on a single error type. This study follows in line with Bruton's (2009) call and van Beuningen et al.'s (2012) and Hartshorn et al.'s (2010)

attempt to retain ecological classroom validity in WCF research by focusing on a range of errors so as to avoid framing WCF as written grammar exercises rather than authentic writing tasks. As Ferris (2010) and Storch (2010) argue, L2 writers and teachers are interested in developing accurate production across linguistic features simultaneously.

A main pedagogical implication of this study is the further support it provides that WCF can be effective in developing students' written grammatical accuracy. Without a clear benefit for overall feedback type in this study (or in the meta-analytic results of Kang and Han, 2015), teachers should be somewhat encouraged to provide either direct or indirect feedback, as both have proven empirically to be at least moderately effective in several studies with continued development in the quality of methodological design. For the time being, for teachers in search of guidance as to which feedback type to provide, it could be best to follow their own preferences and/or those of their students in deciding an appropriate approach.

Can we say then, that there is value in focusing effort on the treatment of particular error types? Examining the target structures in isolation, it appears that error type did play a role in the relative amenability to development in accuracy in this sample considering the variable gains across the different error types and in response to different types of feedback. Results of this study suggest that certain error types may be more amendable to feedback in general (e.g., singular-plural noun agreement, articles) than others (e.g., word forms), although few statistically significant patterns in the data suggest caution in drawing such conclusions. At this point, from a practical perspective, fine-grained analysis of results for individual error types could be helpful for practitioners, particularly if research contexts align with student and classroom characteristics. For instance, I would advise teachers in the context of this Thai university (where general policy is to provide feedback on a wide range of error categories) to consider ignoring

part of speech and run-on sentence errors if they prefer indirect feedback techniques. For those who currently provide direct feedback, exposure to these results may influence them to switch to indirect in view of the descriptive findings. It is also noteworthy that the effects found in this study were based on only five weeks of treatment. If a greater quantity of writings can be expected in more than five weeks of treatment in other writing contexts, the effects could be more pronounced.

In terms of grammatical knowledge, the IEP where this study was conducted has adopted parts of the Grammatical Knowledge Test designed for these students to use as an online diagnostic for students entering level 3 classes. While results of this study may not inform their feedback choices, it will likely be used to help raise awareness of grammatical knowledge and to focus the topics of explicit grammar study in the writing class. Following the data collection period in this study, the Grammatical Knowledge test results were used to help assign explicit grammar exercises for the areas in which students were weakest in their scores. This diagnostic allowed much of the explicit grammar content of the course (built in to the syllabus) to be covered more efficiently in the last three weeks of the course. In programs that teach grammar explicitly in conjunction with writing instruction, such a test could be useful to tailor assignments or instruction.

In making choices about WCF, careful and detailed consideration of context is necessary in terms of a writing program, assignment type, one's own beliefs as a writing teacher, as well as students' perceptions and what resources they bring to a writing classroom. Teachers then need to consider which contextual factors are feasible and reasonable to accommodate, and how best to do so. As we learn more about the role of grammatical knowledge in WCF, it could help inform teachers' strategies in providing a mixed approach in their feedback depending on their

perception of a students' ability or knowledge, the linguistic target, the density of errors and types, and the confidence a teacher has in a student's ability or motivation. Experienced teachers can assess these factors while making decisions about how to provide WCF, and by becoming aware of the relationship between linguistic target, grammatical knowledge, and feedback type, it can better equip practitioners to make more informed choices in a flexible and responsive approach that reflects critical examination of our practices.

### **Limitations of the Study**

Several limitations in the design of this study have been mentioned earlier, and more are highlighted here in terms of the general design and the method of analysis. One clear limitation is the lack of delayed posttest. With only seven weeks available for data collection, a longer treatment was chosen in favor of a delayed posttest, although the posttest was a combination of performance in two writings a week apart with no significant change in performance. Without a delayed posttest of a few weeks though, we cannot be sure if the gains in accuracy would persist. Another limitation in the study design is that success in revision was not measured, which is a necessary first step in the pathway to development. Knowing where students are successful in revision would be helpful contextualize gains or lack thereof in response to different feedback types. This study also ignored a wide range of "nongrammatical" error types (e.g., word choice, prepositions, pragmatic errors) that were prominent in this context. These features did not lend themselves to the measurement of grammatical knowledge in this study, but it would be useful to build evidence regarding the hypothesized treatability of grammatical vs nongrammatical linguistic targets. Another potential limitation in the design was a lack of measure for language complexity (this was attempted in van Beuningen, et al., 2012) to account for possible tradeoff between complexity, accuracy, and fluency in language development. Complexity was ignored

here considering the size of the data set and that no clear complexity measure has been agreed on for particular proficiencies or writing task types (e.g., Norris & Ortega, 2009).

Finally, in terms of the analysis, there is no dispute that correct usage in obligatory context is a more accurate measure of grammatical accuracy than normed error rates. Obligatory use would have been preferable, as it has been employed in several more recent WCF studies, although in each case they have been highly focused (on one or two error types). However, the size of this data set, the wider focus on six error types, and the fact that some error types could not be measured in terms of obligatory use (e.g., run-on sentence errors), error rates were used. In terms of the statistical analysis, there are limitations of bivariate correlations to determine the relationship between the grammatical knowledge and gains in accuracy. The third variable problem, noted by Field (2005) refers to the unmeasured variables that affect results, and high variation in the data suggest several potential moderating variables. This remains a challenge in examining the impact of cognitive factors on the effectiveness of different feedback conditions. Hopefully we can continue to improve on these limitations in future study design.

### **Suggestions for Future Research**

Regarding the debate over the effectiveness of WCF in L2 writing pedagogy, as research shows clearer evidence of limited effectiveness in recent decades (supported by the results of this study), research attention should shift to longitudinal investigations of the persistence of effectiveness. With a better understanding of how the effects of WCF persist, not only in delayed post-tests of a few weeks but in subsequent semesters and with long (semester) breaks between treatments, we could better understand the extent to which effects last months or years following treatments and whether continued feedback in later semesters maintains similar effect or

reinforces prior gains. With so much effort required to provide WCF, curriculum planners would benefit with evidence for optimal durations and sequencing of feedback.

To move research forward that investigates the effects of WCF on development in accuracy, more studies are needed that explore the role of potential moderating variables. With systematic decisions for focusing on linguistic targets, studies in diverse contexts can provide further evidence, especially at the meta-analytic level. The post hoc analysis in this study suggests that grammatical knowledge likely plays a role in WCF, and that grouping learners according to grammatical knowledge for analysis could be fruitful. It could also be that grammatical knowledge interacts with other potential variables (e.g., student preference, motivation) so data on a wider range of potential moderator variables is encouraged. For instance, student surveys of their preferences for feedback type were conducted following the treatment in this study, and will be the subject of further analysis. Posttest data for the Grammatical Knowledge Test was also collected following the treatment (the same test), but was not analyzed for the present study. Analysis of the posttest results could reveal the effect of WCF on development in learners' explicit grammatical knowledge, which may often be an intended result of WCF (outside of the development of accuracy). Another construct related to grammatical knowledge that deserves attention in relation to WCF is learners' meta-awareness of their grammatical knowledge or their strengths and weaknesses in accuracy. Surveys were also collected before and after the treatment in this study that asked students to comment on their strengths and weaknesses in accuracy. A fourth class was provided an alternative form of indirect feedback using color highlights as a code. Analysis of this data can explore the potential value of learners' meta-awareness of their own strengths and weaknesses in terms of how students develop in their written accuracy.

From a methodological standpoint, there are a few considerations from this study that could help guide improved WCF research design. First, with increased attention on linguistic targets moving forward, it may be helpful to reimagine our analysis of error types, i.e., more in line with the sub-categories used in this study to develop the Grammatical Knowledge Test items. Irregular forms, for example, likely behave differently in response to different feedback types, and analyzing the nuances that exist within error categories as they are typically delineated could be revealing. Also, incorporating analysis of textual differences between task types could be useful to account for the differences in task type. Differences in writing topic influence linguistic features in texts, which in turn, would influence patterns in error. Future incorporation of corpus tools to analyze patterns in the use of linguistic features could help control for the influence of task type. For example, a lexical analysis of the ratio between regular and irregular nouns across writing topics (and their relative accuracy rates) could help to more tightly control study design.

## References

- Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research, 16*, 9-37.
- Alderson, J. C., Clapham, C., & Steel, D. (1997). Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching Research, 2*, 93-121.
- Ammar, A., & Spada, N. (2006). One size fits all? Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition, 28*, 543-574.
- Benevento, C., & Storch, N. (2011). Investigating writing development in secondary school learners of French. *Assessing Writing, 16*, 97-110.
- Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning, 29*, 81-103.
- Biber, D, Nekrasova, T and Horn, B (2011) The effectiveness of feedback for L1-English and L2 writing development: A meta-analysis. *TOEFL iBT Re-search Report No. TOEFLiBT-14*. Princeton, NJ: Educational Testing Service.
- Birdsong, D. (1989). *Metalinguistic performance and interlanguage competence*. Berlin, Germany: Springer-Verlag.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*, 102-118.
- Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.
- Bitchener, J, & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*, 409-431.
- Bitchener, J., & Knoch, U. (2010). The contribution of written corrective feedback to



- language development: A ten-month investigation. *Applied Linguistics*, 31, 193-214.
- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14, 191-205.
- Brown, D. (2012). The written corrective feedback debate: Next steps for classroom teachers and practitioners. *TESOL Quarterly*, 46, 861-867
- Bruton, A. (2009). Designing research into the effect of error correction in L2 writing: Not so straightforward. *Journal of Second Language Writing*, 8, 136-140.
- Butler, Y. (2002). Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition*, 24, 451-480.
- Chaney, S. J. (1999). *The effect of error types on error correction and revision*. Sacramento, CA: California State University, Department of English, M. A. thesis.
- 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), 267-296.
- Clapham, C. (2001). The assessment of metalinguistic knowledge. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. Macnamara, & K. O'Loughlin (Eds.), *Experiencing with uncertainty: Essays in honour of Alan Davies* (pp. 31-41). Cambridge, England: Cambridge University Press.
- Correa, M. (2011). Heritage language learners of Spanish: What role does metalinguistic knowledge play in their acquisition of the subjunctive? Selected proceedings of the 13<sup>th</sup> Hispanic Linguistics Symposium, (Ed.) L. A. Ortiz-López (pp. 128-138).
- Cronbach, L. J. (1967). How can instruction be adapted to individual differences. *Learning and individual differences*, 23-39.

- DeKeyser, R. (2007). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (pp. 979-113). Mahwah, NJ: Erlbaum.
- Elder, C. (2009). Validating a test of metalinguistic knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philip and H. Reinders, *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*. Bristol: Multilingual Matters, pp. 113-38.
- Elder, C., Erlam, R., & Philp, J. (2007). Explicit language knowledge and focus on form: Options and obstacles for TESOL teacher trainees. In S. Fotos & H. Nassaji (Eds.), *Form-focused instruction and teacher education: Studies in honour of Rod Ellis* (pp. 225-240). Oxford: Oxford University Press.
- Elder, C., Warren, J., Hajek, J., Manwaring, D., & Davies, A. (1999). Metalinguistic knowledge: How important is it in studying at university? *Australian Review of Applied Linguistics*, 22, 81-95.
- Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning*, 54, 227-275.
- Ellis, R. (2010). A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, 32, 335-349.
- Ellis, R., & Sheen, Y. (2006). Reexamining the role of recasts in second language acquisition. *Studies in Second Language Acquisition*, 28, 575-600.
- Ellis, R., Sheen, Y., Murakami, M., & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context *System*, 36, 353-371.
- Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, 9, 147-171.

- Ferris, D. (1995). Teaching ESL composition students to become independent self-editors. *TESOL Journal*, 4, 18-22.
- Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8, 1-10.
- Ferris, D. (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, 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 K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). Cambridge, U.K.: Cambridge University Press.
- Ferris, D. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, 32, 181-201.
- Ferris, D. (2011). *Treatment of error in second language student writing*. University of Michigan Press.
- Ferris, D., Chaney, S. J., Komura, K., Roberts, B. J., & McKee, S. (2000, March). Perspectives, problems, & practices in treating written error. Colloquium presented at International TESOL Convention, Vancouver, BC.
- Ferris, D., & Roberts, B. J. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10, 161-184.
- Ferris, D., & Hedgcock, J. S. (2004). *Teaching ESL composition: Purpose, process, and practice*. Mahwah, NJ: Lawrence Erlbaum.

- Field, A. (2005). *Discovering statistics using SPSS*. Thousands Oaks, CA: SAGE Publications Inc.
- Gass, S., & Varonis, E. (1994). Input, interaction and second language production. *Studies in Second Language Acquisition*, *16*, 283-302.
- Godfroid, A., Loewen, S., Jung, S., Park, J., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, *37*, 269-297.
- Goldstein, L. (2004). Questions and answers about teacher written commentary and student revision: Teachers and students working together. *Journal of Second Language Writing*, *13*, 443-460.
- Green, P., & Hecht, K. (1992). Implicit and explicit grammar: An empirical study. *Applied Linguistics*, *13*, 168-184.
- Gu nette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, *16*, 40-53.
- Guti rrez, X. (2013). Metalinguistic knowledge, metalingual knowledge, and proficiency in L2 Spanish. *Language Awareness*, *22*, 176-191.
- Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research*, *2*, 1-23.
- Hartshorn, J., Evans, N. W., Merrill, P. F., Sudweeks, R. R., Strong-Krause, D., & Anderson, N. J. (2010). Effects of dynamic corrective feedback on ESL writing accuracy. *TESOL Quarterly*, *44*, 84-109.
- Hendrickson, J. M. (1980). The treatment of error in written work. *Modern Language Journal*. *64*, 216-221.

- Hu, G. (2002). Psychological constraints on the utility of metalinguistic knowledge in second language production. *Studies in Second Language Acquisition*, 24, 347–386.
- Hyland, K., & Hyland, F. (2006). *Feedback in Second Language Writing: Contexts and Issues*. New York, NY: Cambridge University Press.
- Jeon, E. H. (2015). Multiple Regression. In *Advancing quantitative methods in second language research*. (Ed. by L. Plonsky), pp. 131-158. New York: Routledge.
- Kang, E. & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal*, 99, 1-18
- Komura, K. (1999). Student response to error correction in ESL classrooms. MA thesis, Department of English, California State University, Sacramento.
- Lalande, J. F. II (1982). Reducing composition errors: An experiment. *Modern Language Journal*, 66, 140-149.
- Larsen-Freeman, D. (2009). Teaching and Testing Grammar. In *The handbook of language teaching*, (Eds.) M. Long. & C. Doughty. (pp. 518-542).
- Larson-Hall, J. (2009). *A guide to doing statistics in second language research using SPSS*. Routledge.
- Lee, I. (1997). ESL learners' performance in error correction in writing: Some implications for college-level teaching. *System*, 25, 465–477.
- Li, S. (2010). *Corrective feedback in perspective: The interface between feedback type, proficiency, the choice of target structure, and learners' individual differences in working memory and language analytic ability*. Unpublished doctoral dissertation. Michigan State University.

- 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, 634-654.
- Liu, Q. (2016). *Effectiveness of coded corrective feedback in the development of linguistic accuracy in L2 writing: Impact of error type and learner attitudes*. Unpublished doctoral dissertation. Northern Arizona University.
- Liu, Q. & Brown, D. (2015). A methodological synthesis of research on the effectiveness of corrective feedback in L2 writing. *The Journal of Second Language Writing*, 30, 66-81.
- Lyster, R., & Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32, 265-302.
- Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching*, 46, 1-40.
- Mackey, A. (2006). Feedback, noticing, and instructed second language learning. *Applied Linguistics*, 27, 405-430.
- Masny, D. (1987). The role of language and cognition in second language metalinguistic awareness. In J. Lantolf & A. Labarca (Eds.), *Research in second language learning: Focus on the classroom* (pp. 59–73). Norwood, NJ: Ablex.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language learning*, 50, 417–528.
- Norris, J. M., & Ortega, L. (2006). The value and practice of research synthesis for language learning and teaching. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 3–50). Philadelphia: Benjamins.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in

- instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555-578.
- Park, E. S. Song, S., & Shin, Y. K. (2016). To what extent do learners benefit from indirect written corrective feedback? A study targeting learners of different proficiency and heritage language status. *Language Teaching Research*, 20, 678-699.
- Pienemann, M. (1998). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins.
- Plonsky, L. (2015). *Advancing quantitative methods in second language research*. (Ed. by L. Plonsky), New York: Routledge.
- Plonsky, L., & Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have different results). *Second Language Research*, 31, 267-278.
- Plonsky, L., & Oswald, F. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878-912.
- Pongsiriwet, C. (2001). Relationships among grammatical accuracy, discourse features, and the quality of second language writing: The case of Thai EFL learners. Unpublished doctorate dissertation, West Virginia University. Morgantown, WV.
- Purpura, J. (2004). *Assessing grammar*. Cambridge: Cambridge University Press.
- Purpura, J. (2006). Issues and challenges in measuring SLA. Paper presented at the American Association for Applied Linguistics Conference, June, Montreal.
- Purpura, J. E. (2013). Assessing grammar. In A. J. Kunnan (Ed.), *The Companion to language assessment* (pp. 100-124). Wiley Blackwell.
- Rahimi, M. (2014). The role of individual differences in L2 learners’ retention of written corrective feedback. *Journal of Responding to Writing*. Retrieved from

<http://journalrw.org/index.php/jrw/index>.

- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning, 63*, 595–626.
- Riazantseva A. (2012). Outcome measure of L2 writing as a mediator of the effects of corrective feedback on students' ability to write accurately. *System, 40*, 421-430.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly, 20*, 83-95.
- Robinson, P. (1997). Individual differences and fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*, 45-99.
- Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*, 27-57.
- Roehr, K., & Gánem-Gutiérrez, G.A. (2009). The status of metalinguistic knowledge in instructed adult L2 learning. *Language Awareness 18*, 165-181.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.
- Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for the acquisition of L2 grammar. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 133–164). Amsterdam: Benjamins.
- Sampson A. (2012). Coded and uncoded error feedback: Effects on error frequencies in adult Columbian EFL learners' writing. *System, 40*, 494-504.
- Serrano, R. (2011). From metalinguistic instruction to metalinguistic knowledge, and from metalinguistic knowledge to performance in error correction and oral production tasks. *Language Awareness, 20*, 1-16.



- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge: Cambridge University Press.
- Semke, H. (1984). The effects of the red pen. *Foreign Language Annals*, 17, 195–202.
- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41, 255-283.
- Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, 32, 203–234.
- Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, 37, 556--569.
- Sorace, A. (1985). Metalinguistic knowledge and language use in acquisition- poor environments. *Applied Linguistics*, 6, 239–254.
- Spada, N., Jessop, L., Tomita, Y., Suzuki, W. & Valeo, A. (2014). Isolated and integrated form-focused instruction: Effects on different types of L2 knowledge. *Language Teaching Research*, 1-21.
- Storch, N. (2010). Critical feedback on written corrective feedback research. *International Journal of English Studies*, 10, 29–46.
- Storch, N. & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing. *Studies in Second Language Acquisition*, 32, 303-334.
- Tellier, A. & Roehr-Brackin, K. (2013). Metalinguistic awareness in children with different language learning experience. *EUROSLA Yearbook*, 13, 81-108.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, 327–369.

- Truscott, J. (1999). The case for “The case against grammar correction in L2 writing classes”: A response to Ferris. *Journal of Second Language Writing*, 8, 111-122.
- Truscott, J. (2007). The effect of error correction on learners’ ability to write accurately. *Journal of Second Language Writing*, 16, 255–272.
- Truscott, J., & Hsu, A. Y. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, 17, 292-305.
- Van Beuningen, C. (2010). Corrective feedback in L2 writing: Theoretical perspectives, empirical insights, and future directions. *International Journal of English Studies*, 10, 1–27.
- Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners’ written accuracy. *ITL International Journal of Applied Linguistics*, 156, 279–296
- Van Beuningen, C. G., de Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning*, 62, 1-41.
- Van Lier, L. (1988). *The classroom and the language learner: Ethnography and second language classroom research*. New York: Longman.
- VanPatten, B. (2007). Input processing in adult second language acquisition. In B. VanPatten & J. Williams, *Theories in second language acquisition* (pp. 115-135). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wesche, M. (1981). Language aptitude in measures in streaming, matching students with methods, and diagnosis of learning problems. In K. Killier (Ed.), *Individual differences and universals in language learning aptitude* (pp. 119-154). Rowley, MA: Newbury

House.

Wistener, B. (2014). *Effects of metalinguistic knowledge and language aptitude on second language learning*. Unpublished doctoral dissertation, Temple University. Philadelphia, PA.

Xu, C. (2009). Overgeneralization from a narrow focus: A response to Ellis et al. (2008) and Bitchener (2008). *Journal of Second Language Writing, 18*, 270– 275.

Yilmaz, Y. (2013). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics, 34*, 344-368.

Zietek, A. A. & Roehr, K. (2011). Metalinguistic knowledge and cognitive style in Polish classroom learners of English. *System, 39*, 417-426.

Appendix A

Grammatical Knowledge Test and Answer Key

**Part 1: Error Correction and Rule Explanation**

1. **Currently, over 100 million babies were born every year.**

(1) Correct the error:

**are**

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>Verb Tense</b>	(past when present required)
1 point (reasoning)	Some reference to the timing of the action: occurring regularly / habitually (e.g., habit / fact / ongoing / regular / continuously)
1 point (language)	verb / present / tense / habitual action / not past / action
2 points (full credit)	‘Currently’ indicates present action and ‘every year’ indicates an habitual action, both of which requires the present tense for “be”.

2. **The program help students to pay for their books.**

(1) Correct the error:

**helps**

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>S-V Agreement</b>	
1 point (reasoning)	Some reference to the subject (or “program”)
1 point (language)	singular / third-person / agreement / conjugation / verb ending
2 points (full credit)	The subject is singular / third-person so the verb needs an –s to agree.

3. **In the last ten years, climate change becomes the most significant issue around the world.**

(1) Correct the error:

**has become OR became**

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>Verb Tense</b>	(perfect aspect)
1 point (reasoning)	Some reference to the timing of the action (e.g., last ten years / past / not present anymore / not finished / began / continues / until now / not stopped / ongoing).
1 point (terms)	(present) perfect / aspect / auxiliary / not present / past / tense / verb 2 / verb 3
2 points (full credit)	This verb requires the present perfect tense because the time when the action occurred began in the past and continues to the present.

4. **One problem with limited resources is that people don't have enough the water to drink across the world.**

**enough water (no article)**

(1) Correct the error:

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>Article</b>	(definite used with uncountable noun)
1 point (reasoning)	Some reference to the type/quality of the word "water" (e.g., noncount / uncountable / plural / not specific) or that the preceding word "enough" isn't followed by an article.
1 point (terms)	uncountable / noncount / abstract noun / article / determiner / specifier / noun
2 points (full credit)	Don't use an article with an uncountable noun.

5. **We need to find a peace solution to prevent war.**

**peaceful**

(1) Correct the error:

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>Part of Speech</b>	(Noun when adj. required)
1 point (reasoning)	Some reference to the choice of the type / position of the word, or the requirements regarding the grammatical environment of the word.
1 point (terms)	noun / adjective / part of speech / word form / word family / before a noun / modifies / direct object
2 points (full credit)	The wrong word form / family / p.o.s. is used. An adjective is needed instead of the noun because it describes the noun "solution".

6. **There are multiple form of transportation that people can choose from.**

(1) Correct the error:

**forms**

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>S-P Agreement</b>	
1 point (reasoning)	Some reference to the reason (e.g., pointing to "multiple" or "are" requiring something different).
1 point (terms)	agreement / (countable) noun / plural / quantifier / singular / many / more than one
2 points (full credit)	The noun is countable, so after 'multiple' use the plural form.

7. It teaches students everything about business, it emphasizes how to start your own business

Start a new sentence  
OR any suitable conjunction

(1) Correct the error:

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>Run-On</b>	
1 point (reasoning)	Reference to the idea of needing something (e.g., period / ‘.’ / ‘and’ / ‘because’ / ; ) to join or separate, or reference to the relationship between the two parts of the sentence / reference to repeating subject
1 point (terms)	two sentences / two independent clauses / two complete thoughts / run-on / subordinator / adverbial / conjunction / connection / transition / punctuation
2 points (full credit)	It is a run-on because two sentences / thoughts / clauses are connected without appropriate subordinator / semi-colon / conjunction.

8. That is very difficult situation for Thai people.

(1) Correct the error:

is a very

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>Article</b>	(missing indefinite)
1 point (reasoning)	Reference to the type of noun/word/object (e.g., countable, specific, only one, singular)
1 point (terms)	indefinite / article / determiner
2 points (full credit)	You need an article / determiner before a countable noun.

9. The two security systems that we have is strong enough to keep us safe.

(1) Correct the error:

are

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>S-V Agreement</b>	
1 point	Some reference to the subject or its nature (e.g., pair / more than one / -s on “systems”)
1 point	subject / plural / agreement / noun / verb
2 points (full credit)	The subject is plural so the verb needs to agree with the noun.

10. In general, people don't know the dangerous of this problem.

(1) Correct the error:

**dangers OR danger**

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>Part of Speech</b>	(Adj. when noun required)
1 point	Some reference to the choice of the type or position of the target word, or the requirements regarding the grammatical environment.
1 point	noun / adjective / part of speech / word form / word family / modify / modifies
2 points (full credit)	The noun form of the word is needed / follows article & precedes PP noun modifier / direct object

11. Thus, when these children growing up, they will become responsible.

(1) Correct the error:

**grow OR are growing OR have grown  
OR are grown**

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>Verb Tense</b>	(attempted progressive aspect when simple necessary)
1 point	Reference to when the action is occurring / Reference to the main verb / Reference to the construction error of auxiliary 'be' + -ing ending
1 point	tense / aspect / simple / present / perfect / conditional / verb / if clause / verb 1
2 points (full credit)	The present tense is needed because the tense is already marked with "will become".

12. Some classes have better teachers. For example, teacher in my math class is teaching well.

(1) Correct the error:

**the teacher**

(2) Explain what is wrong with the sentence. What grammar rule can explain your correction?

<b>Article</b>	(missing definite)
1 point	Reference to the nature/type of the word/noun that requires something (e.g., countable, one teacher, singular, specific)
1 point	definite / article / determiner / noun (if relevant) / specific
2 points (full credit)	You need a article before a specific noun. Use the definite article here because it refers to a specific teacher.

## Part 2: Error Correction

1. The students should study all subject in their first year.

**subjects**

2. It is the first thing I thought about each day when I wake up.

**I think / am thinking**

3. The most convenience way to travel to university is by taxi, except during awful traffic.

**convenient**

4. Everything that appear in the homework will be covered on the test.

**appears** (also: **had appeared / appeared**)

5. They will be worried to ask and will not talking with their parents about the problem.

**talk**

6. Parents should be role models for their children, they will learn more positive things from their parents.

**children and they** (or any suitable conjunction) OR **children. They...**

7. Students need to work hard and take the time every day to study.

**take time** OR **take their time**

8. Some people believe those girls who have a baby before they get married is bad people.

**are**

9. It is safer at home so staying at home is better place for kids to stay.

**is a better** OR **the best**

10. That type of school has worst environment for kids.

**has the worst**

11. Their responsibilities at work change since the new director is leading the program.

**have changed** OR **changed** OR **have been changed** OR **have been changing**

12. This skill set is required for successful in the job market after graduation.

**success** OR **to succeed** OR **to be successful** OR **for succeeding** OR **for being successful**



### Part 3: Multiple Choice

1. Other types of student will learn faster.
  - A. Subject-Verb Agreement
  - B. Part of Speech (Word Family)
  - C. Singular-Plural Noun Agreement**
  - D. No Error
  
3. Since this class began this become a very common problem.
  - A. Run-On Sentence
  - B. Verb Tense / Aspect**
  - C. Article
  - D. No Error
  
5. Helping them is not a problem in Thailand because people respect elders.
  - A. Verb Tense/ Aspect
  - B. Run-On Sentence
  - C. Article
  - D. No Error**
  
7. Therefore, my school encourage students to succeed in their lives.
  - A. Run-On Sentence
  - B. Subject-Verb Agreement**
  - C. Singular-Plural Agreement
  - D. No Error
  
2. Parents can indicate what is good thing to do for their children.
  - A. Verb Tense / Aspect
  - B. Article**
  - C. Subject-Verb Agreement
  - D. No Error
  
4. If they failed when the new results are released soon, they will find a solution by themselves.
  - A. Part of Speech (Word Family)
  - B. Subject-Verb Agreement
  - C. Verb Tense / Aspect**
  - D. No Error
  
6. We need a logic solution to this simple problem.
  - A. Part of Speech (Word Family)**
  - B. Run-On Sentence
  - C. Subject-Verb Agreement
  - D. No Error
  
8. Most parents want their kids to have better future.
  - A. Subject-Verb Agreement
  - B. Run-On Sentence
  - C. Article**
  - D. No Error

9. They do not strictly follow the rules like we did in high school.

- A. Part of Speech (Word Family)
- B. Run-On Sentence
- C. Verb Tense/ Aspect
- D. No Error**

11. I hope to earn good grade in this class.

- A. Article**
- B. Verb Tense/ Aspect
- C. Part of Speech
- D. No Error

13. Therefore, teaching the true is the best thing parents can do for their children.

- A. Subject-Verb Agreement
- B. Verb Tense / Aspect
- C. Part of Speech (Word Family)**
- D. No Error

15. No one is too young to decide what they want, this is also true for children too.

- A. Singular-Plural Noun Agreement
- B. Run-On Sentence**
- C. Verb Tense/ Aspect
- D. No Error

17. If someone gets pregnancy as a teenager it could cause difficulty in their family.

- A. Subject-Verb Agreement
- B. Article
- C. Part of Speech (Word Family)**
- D. No Error

10. This will be the big problem if we do not solve it now.

- A. Run-On Sentence
- B. Article**
- C. Singular-Plural Noun Agreement
- D. No Error

12. Students still do not understanding about the issue properly.

- A. Part of Speech (Word Family)
- B. Singular-Plural Noun Agreement
- C. Verb Tense / Aspect**
- D. No Error

14. Some people think new teacher is very kind.

- A. Article**
- B. Subject-Verb Agreement
- C. Run-On Sentence
- D. No Error

16. Parents are important for children when they are young because they were in control of their children.

- A. Verb Tense / Aspect**
- B. Part of Speech (Word Form)
- C. Run-On Sentence
- D. No Error

18. Our first subject is at 8am and last subject ends at 3pm.

- A. Part of Speech (Word Family)
- B. Article**
- C. Subject-Verb Agreement
- D. No Error

19. However, as the world moving forward we need more solutions.

- A. Article
- B. Run-On Sentence
- C. Verb Tense/ Aspect**
- D. No Error

21. We have been trying this solution for three years but each time it failing to help us.

- A. Singular-Plural Noun Agreement
- B. Part of Speech (Word Family)
- C. Verb Tense / Aspect**
- D. No Error

23. I have a friend whose parents do not have time to take care of him.

- A. Singular-Plural Noun Agreement
- B. Part of Speech (Word Family)
- C. Verb Tense/Aspect
- D. No Error**

25. You can avoid traffic by using sky train in Bangkok.

- A. Article**
- B. Subject-Verb Agreement
- C. Singular-Plural Noun Agreement
- D. No Error

27. Students of all ages around the world do not like to do the homework.

- A. Part of Speech (Word Family)
- B. Verb Tense / Aspect
- C. Article**
- D. No Error

20. This way only tell them it is wrong without showing the reason why.

- A. Subject-Verb Agreement**
- B. Article
- C. Part of Speech (Word Family)
- D. No Error

22. It won't happen again because that was an accident mistake.

- A. Part of Speech (Word Family)**
- B. Verb Tense/ Aspect
- C. Run-On Sentence
- D. No Error

24. Children will gain a lot of knowledge, maybe they can learn from friends or from traveling too.

- A. Subject-Verb Agreement
- B. Singular-Plural Noun Agreement
- C. Run-On Sentence**
- D. No Error

26. When children is restricted by their parents they will grow up slower than children with freedom.

- A. Run-On Sentence
- B. Subject-Verb Agreement**
- C. Part of Speech (Word Family)
- D. No Error

28. Since I was a boy I grow 60 cm.

- A. Article
- B. Subject-Verb Agreement
- C. Verb Tense/Aspect**
- D. No Error

29. Children will not develop the same society skills if they always stay at home.

- A. **Part of Speech (Word Family)**
- B. Subject-Verb Agreement
- C. Run-On Sentence
- D. No Error

31. Today the debate begin about the best solution to this problem.

- A. Run-On Sentence
- B. Singular-Plural Noun Agreement
- C. **Subject-Verb Agreement**
- D. No Error

33. Furthermore, teachers can encourage students to do a lot of homework.

- A. Singular-Plural Noun Agreement
- B. Run-On Sentence
- C. Subject-Verb Agreement
- D. **No Error**

35. This issue is a useful thing to talk about in the society in every country.

- A. **Article**
- B. Run-On Sentence
- C. Verb Tense/ Aspect
- D. No Error

37. It is better to provide children many optional and let them make their own choice.

- A. Subject-Verb Agreement
- B. Article
- C. **Part of Speech (Word Family)**
- D. No Error

30. Most parent can believe in the quality of the education at this school.

- A. Part of Speech (Word Family)
- B. **Singular-Plural Noun Agreement**
- C. Verb Tense / Aspect
- D. No Error

32. If people is allowed more choices in life, they will be happier .

- A. **Subject-Verb Agreement**
- B. Part of Speech (Word Family)
- C. Verb Tense / Aspect
- D. No Error

34. Lots of things are changing every day as more technology was invented each year.

- A. Run-On Sentence
- B. Subject-Verb Agreement
- C. **Verb Tense / Aspect**
- D. No Error

36. If parents give their children too much freedom they will risk having bad behavior, it can be true in some cases.

- A. Part of Speech (Word Family)
- B. **Run-On Sentence**
- C. Subject-Verb Agreement
- D. No Error

38. We can find it in media such as television, magazine, and newspapers.

- A. Run-On Sentence
- B. **Singular-Plural Noun Agreement**
- C. Part of Speech (Word Family)
- D. No Error

39. I never sing that song when I lived in Australia.

- A. Run-On Sentence
- B. Verb Tense/ Aspect**
- C. Article
- D. No Error

40. Some parents choose to educate their children at home because there is so many benefits.

- A. Part of Speech (Word Family)
- B. Subject-Verb Agreement**
- C. Verb Tense/Aspect
- D. No Error

## Appendix B

### Table of Specifications for Grammatical Knowledge Test

Type of test: Proficiency; Response Type: Supply/ Identify

Scoring:

Guided Error Correction (Procedural/Declarative): Right/wrong, 1 point correct, 0 incorrect

Metalinguistic description (Declarative): Analytic Scale 0-2 points

Error Type Recognition/Identification (Declarative): 1 point correct, 0 points incorrect

Error Category	Part 1: Metalinguistic Knowledge (Declarative Knowledge)	Part 2: Error Correction (Procedural/Declarative Knowledge)	Part 3: Error Type Recognition/ Identification (Procedural/Declarative Knowledge)	# of tasks	# of points	% of points
Singular-plural agreement (only 1 type)	0-2 pts #6	0-2 pts part 1 #6, part 2 #1	0-3 pts #1, #30, #38	6	7	16.7%
Articles (Type 1)	0-2 pts #8	0-2 pts part 1 #8, part 2 #2	0-3 pts #2, #8, #11	6	7	5.6%
Articles (Type 2)	0-2 pts #12	0-2 pts part 1 #12, part 2 #10	0-3 pts #14, #18, #25	6	7	5.6%
Articles (Type 3)	0-2 pts #4	0-2 pts part 1 #4, part 2 #7	0-3 pts #27, #34, #35	6	7	5.6%
Subject-verb agreement (Type 1)	0-2 pts #2	0-2 pts part 1 #2, part 2 #4	0-3 pts #3, #7, #20	6	7	8.35%
Subject-verb agreement (Type 2)	0-2 pts #9	0-2 pts part 1 #9, part 2 #8	0-3 pts #26, #32, #40	6	7	8.35%
Verb Tense / Aspect (Type 1)	0-2 pts #1	0-2 pts part 1 #1, part 2 #9	0-3 pts #4, #10, #12	6	7	5.6%
Verb Tense / Aspect (Type 2)	0-2 pts #11	0-2 pts part 1 #11, part 2 #5	0-3 pts #16, #19, #21	6	7	5.6%
Verb Tense / Aspect (Type 3)	0-2 pts #3	0-2 pts part 1 #3, part 2 #11	0-3 pts #28, #31, #39	6	7	5.6%
Part of Speech (Type 1)	0-2 pts #5	0-2 pts part 1 #5, part 2 #3	0-3 pts #6, #13, #17	6	7	8.35%
Part of Speech (Type 2)	0-2 pts #10	0-2 pts part 1 #10, part 2 #12	0-3 pts #22, #29, #37	6	7	8.35%
Run-on sentences (only 1 type)	0-2 pts #7	0-2 pts part 1 #7, part 2 #6	0-3 pts #15, #24, #36	6	7	16.7%
No Error (distractors)			#5, #9, #23, #33	4		
# of tasks	12	24	36	72		
# of points	24	24	36		84	

Appendix C

Training Materials for Coded Feedback

**Part 1.** Each sentence has a grammar error. With a partner, make the corrections below each grammar error.

SV  
1. My grandmother like to travel  
\_\_\_\_\_

SP  
2. I have two test tomorrow.  
\_\_\_\_\_

VT  
3. When I was a high school student, I play the piano.  
\_\_\_\_\_

Art  
4. I want to get job in business.  
\_\_\_\_\_

POS  
5. I strong disagree with that idea.  
\_\_\_\_\_

RO  
6. Kids can make new friends at school or playing sports, it can happen anywhere.  
\_\_\_\_\_

SV  
7. Many people is afraid of challenges in life.  
\_\_\_\_\_

SP  
8. She has given me a lot of advices.  
\_\_\_\_\_

VT  
9. Since I was 5 years old I go to Laos four times.  
\_\_\_\_\_

Art  
10. He hopes to get high score on the TOEFL test to study abroad.

**Grammar Key**

SP = Singular-Plural Noun Agreement

Art = Article choice

SV = Subject-Verb Agreement

POS = Part of Speech / Word Family

VT = Verb Tense / Aspect

RO = Run-On Sentence

**Part 2:** Next, imagine this is your writing and read the comments from your teacher about grammar. Follow the comments and use the Grammar Key to make the corrections on this Google document.

Every **children[SP]** in this world have rights of basic education. Today there are many way[SP] to be educated and some people prefer to keep their children at home for schooling, while most parents prefer to send their children to school outside the home. Some parents feel that if **[AE]** child **stay[SV]** home for education then they will not have enough social contact and can not make many **friends, parents[RO]** may not know all **[AE]** information that their children need to learn. Others argue that homeschooling **allow[SV]** parents to make sure their children's quality of education is **strongly[WF]** and they can be the best teachers for their children at home. I **agreed[VT]** with several arguments in support of homeschooling should convince parents to teach their own children.

### Key for Error Types

**SV = Subject-Verb Agreement** - It means the subject and the verb of the sentence need to match grammatically. If the subject is *he, she, or it*, you need to add **-s** to the end of the verb (or **-es**).

**SP = Singular-Plural Noun Agreement** - The noun has to change to show the number of the noun grammatically. Is there only one noun or many? Is the noun countable or uncountable? Do you need to add **-s** to the noun?

**VT = Verb Tense / Aspect** - There is an error in verb tense or aspect. Consider which form of verb you need (present, past, future, progressive, perfect).

**AE = Article choice** - The article choice is incorrect, so you need either an article (*a, the, or an*), or you need to delete the article, or choose a different article. It could help to consider if the noun is specific or general or if it is the first time the noun was introduced.

**WF = Word Family / Part of Speech** - The choice in word form (a so-called part of speech). This means you need to choose the *noun, verb, adjective, or adverb* form of the word.

**RO = Run-on Sentence** - The sentence is a run-on sentence, which means the text is too long grammatically. You probably need to cut the sentence into two sentences using punctuation.



## Appendix D

### Grammar Key for Students (Indirect coded feedback group)

**SP = Singular-Plural Agreement** → It means the subject and the verb of the sentence need to match grammatically. The noun has to change to show the number grammatically.  
(Is there only one noun or many? Is the noun countable or not? Do you need "s" to show many? If the subject is *he*, *she*, or *it*, you need to add -s to the end of the verb (or *-es*).

*Example:* I have two favorite kind of music.

**Art = Article choice** → You need to use an article ("a", "the", or "an"). Or, you need to choose a different article. (This is one of the most difficult grammar points to master but it could help think: Is the noun specific or general? Is it the first time the noun was introduced?)

*Example:* I want to get high score on the TOEFL test.

**SV = Subject-Verb Agreement** → It means the subject and the verb of a sentence need to match grammatically. If the subject is a person, "He", "She" or "It", you need to add "s" or "es" to the end of the verb.

*Example:* My grandmother like to travel with us on holidays.

**VT = Verb Tense/ Aspect** → There is an error in verb tense or aspect. Consider which form of verb you need (present, past, future, progressive, perfect).

(Present----Progressive-----Past-----Future ----- Perfect)

*Example:* When I was a high school student, I play the piano.

**WF = Word Family (Part of Speech)** → This means you need a different word family, for example 1) noun, 2) verb, 3) adjective or 4) adverb. You need to choose one of the other forms.

*Example:* Both of these choices have differently advantages.

**R.O. = Run-On Sentence** → This means the sentence is too long grammatically. You need to break it apart into two or more sentences.

*Example:* Most people face traffic jams every day, most people are on the road.

## Appendix E

### IRB Permission



Institutional Review Board for the  
Protection of Human Subjects in Research

Northern Arizona University  
PO Box 4087  
Flagstaff, AZ 86011-4087

928-523-4340  
928-523-1075 fax  
[www.research.nau.edu/vpr/IRB](http://www.research.nau.edu/vpr/IRB)

**To:** Daniel Brown, BA, MA  
**From:** University of Arizona Irb  
**Approval Date:** January 13, 2016

**Project:** The Interaction between Grammatical Knowledge and Explicitness in L2 Written Corrective Feedback  
**Project Number:** 855324-1

**Review Category/ies:** *Exempt Approval 45 CFR 46.101(b)(1)(ii): Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.*

Your research protocol has been approved by the Human Subjects Committee/Institutional Review Board (IRB) at NAU under the category of EXEMPT. This category means that your IRB approval for this project does not have an expiration date, so periodic renewal of approval is not necessary unless there are changes in your project that affect the status.

If your project **changes** in any way, you must file a Research Amendment form available at <https://www.research.nau.edu/compliance/irb/forms.aspx> PRIOR TO implementing any changes. You may not implement the changes until you have written approval for the change from the IRB, unless the change is necessary to eliminate immediate hazards to participants. Failure to do so will result in noncompliance and possible suspension or termination of your research project.

Any unanticipated problems or unexpected **adverse events** must be reported to the IRB within 5 business days (within 24 hours for serious adverse events) of your becoming aware of the event by filling out an Adverse Reaction or Event Reporting form (also available at website above).

As you conduct your research, please remember that:

1. Participants are volunteers or are involved in regular educational programs; they are free to withdraw from the research at any time without penalty.
2. Unless you are using existing data, Participants must be informed of the research project through written or oral explanation and must sign or approve electronically or verbally an informed consent form (for minors and children the parent or guardian must sign).
3. Unless the participants agreed to an alternative arrangement, the participants' anonymity and confidentiality must be protected. They should not be able to be identified through the responses. The presentation of the data should not put them at risk of any negative consequences. Access to the data is specified and restricted by the researcher and the department.