WRITING IN A TASK-BASED INDIVIDUALIZED CURRICULUM: EFFECTIVENESS OF DIRECT AND INDIRECT WRITTEN CORRECTIVE FEEDBACK

A Dissertation submitted to the Faculty of the Graduate School of Arts and Sciences Of Georgetown University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics

By

Xue Ma, M.S.

Washington, DC

May 11, 2020
WRITING IN A TASK-BASED INDIVIDUALIZED CURRICULUM: EFFECTIVENESS OF DIRECT AND INDIRECT WRITTEN CORRECTIVE FEEDBACK

Xue Ma, M.S.

Thesis Advisors: Alison Mackey, Ph.D. and Ronald Leow, Ph.D.

ABSTRACT

The effects of written corrective feedback have been extensively investigated for various domains of instructed second language acquisition with many studies demonstrating clear benefits (e.g. Riazi, Shi & Haggerty, 2018). However, there are still many unresolved questions about how written corrective feedback is best applied in classroom contexts (e.g., Bitchener, 2008; Chandler, 2003; Ferris, 1995; Benson & DeKeyser, 2018). Studies have found conflicting results in terms of what types of written corrective feedback may be more effective and lead to greater language learning (e.g., Bitchener, 2018; Ferris, 2006; Hartshorn et al., 2010; Manchón, 2011). There is a clear need for research targeting how learners process different types of written corrective feedback in order to address why certain types are more effective and how these feedback types are shaped by different learning contexts (Caras, 2019; Leow, in press; Manchón, 2018; Suzuki, 2017).

Building on previous research, this study investigated the processing and effectiveness of unfocused direct written corrective feedback and unfocused indirect written corrective feedback with metalinguistic codes and examples. The study was conducted in a Chinese foreign language course with thirty-eight L2 Chinese beginners randomly divided into the two feedback conditions. Thirty-eight participants wrote three compositions over four weeks in the task-based curriculum. The study investigates two different grammatical targets, the particle le and classifiers, to explore the effects of linguistic target on written corrective efficacy. Think-aloud protocols and quantitative analyses of accuracy scores were triangulated to assess participants’ processing of written corrective feedback and their performances.
The study found that the indirect-feedback-with-metalinguistic-explanations group exhibited greater depth of processing overall when compared to the direct written corrective feedback group. The indirect-feedback-with-metalinguistic-explanations group demonstrated significantly higher mean accuracy in the delayed posttest for the patterned, rules-based target of the particle le. Think-aloud responses and focus group data suggest that the problem-solving challenge inherent in indirect feedback with metalinguistic explanations may have caused students to exert themselves more and was connected to achieving awareness at the level of understanding (Schmidt, 1990), while direct written corrective feedback was frequently only processed at the level of noticing (Leow, in press). This study provides insights into the effects of direct written corrective feedback versus indirect feedback with metalinguistic explanations. The dissertation concludes with implications for classroom practices and future research.

Key words: written corrective feedback, task-based language teaching, Mandarin Chinese
ACKNOWLEDGMENTS

I would like to express special appreciation and thanks to my advisors, Professor Alison Mackey and Professor Ron Leow. Professor Mackey has been an outstanding mentor for me. Thank you for encouraging me and pushing me to grow as a good researcher, for all your support in my career, and for all the practical advice for a new mom! Professor Ron Leow, thank you for all your encouragement, humor, and guidance on my dissertation! I would also like to express huge thanks to my committee members, Professor Lourdes Ortega and Professor Lara Bryfonski. Thank you for listening patiently to my ideas and working hard to support my research. I am really lucky to have completed my doctoral thesis in such a supportive and positive environment.

Completing this work would have been impossible without the support and help provided by the distance learning program and community at the hosting institution for this study. Special thanks to my manager Kristen, Chinese mentors Shuxiang, Lily, Yutong, and all the students who participated in this study. I am indebted to you for all of your help and support.

A huge thank you to my caring friends at Georgetown who provided perspective and comradery as we climbed the Ph.D. program ladder. Special thanks to Linxi Zhang and Alison Caras for all the deep discussions and feedback on my research design and coding. Di Qi, Viggo, Yiran, Francesca, Ayşenur, thank you for always giving me the positive energy needed to keep going!

Finally, to my family, I would like to thank my parents Xinxia Wu and Longhai Ma, who experienced all the ups and downs of my research. Thank you too to the Hanson family for their support and encouragement. I must express my gratitude to my partner, Mark Hanson, for supporting me through my years at Georgetown—I am still amazed by his “willingness” to proofread countless pages of linguistic papers over the years. I feel like I should give him half of my Ph.D. To my beloved son Thomas Hanson, thank you for being such a great boy and always cheering me up!
# TABLE OF CONTENTS

**CHAPTER I: INTRODUCTION** .................................................................................................................. 1

1.1 Statement of the Problem .................................................................................................................. 1

**CHAPTER II: LITERATURE REVIEW** .................................................................................................. 5

2.1 Task-Based Language Teaching ....................................................................................................... 5

2.2 Effectiveness of Different Written Corrective Feedback Types ...................................................... 10

2.2.1 Focused vs. Unfocused Written Corrective Feedback ................................................................. 15

2.2.2 Focused Written Corrective Feedback ......................................................................................... 20

2.2.3 Unfocused Written Corrective Feedback ..................................................................................... 27

2.2.4 Summary ....................................................................................................................................... 34

2.3 Processing Written Corrective Feedback .......................................................................................... 35

2.3.1 Theoretical Framework for Processing Feedback ........................................................................... 36

2.3.2 Procedures for Assessing Processing ............................................................................................ 41

2.3.3 Process-oriented Studies of Written Corrective Feedback ........................................................ 47

2.3.4 Summary ....................................................................................................................................... 52

2.4 Feedback Engagement and Teacher Perspectives on Written Corrective Feedback ....................... 56

2.5 Current Study .................................................................................................................................... 62

2.5.1 Research Questions ...................................................................................................................... 63

**CHAPTER III: METHODOLOGY** ....................................................................................................... 64

3.1 Participants ......................................................................................................................................... 64

3.2 Context ............................................................................................................................................... 67
4.2.3 Depth of Processing Over Time........................................................................... 115
4.2.4 Summary of Research Question 2 Results.......................................................... 121
4.3 Research Question 3 .............................................................................................. 122
  4.3.1 Student Engagement......................................................................................... 122
  4.3.2 Summary of Research Question Three Results............................................... 130
4.4 Research Question 4 .............................................................................................. 130
  4.4.1 Summary of Research Question Four .............................................................. 133

CHAPTER V: DISCUSSION .......................................................................................... 134
  5.1 Research Question 1: What is the Effect of Different Types of WCF (Direct vs. Indirect with Metalinguistic Explanations) on Low-proficiency L2 Chinese Writers’ Accuracy with Le and Classifiers? .......................................................... 134
  5.2 Research Question 2: How do Low-proficiency L2 Chinese Writers Process the Two Different Types of WCF (Direct vs. Indirect with Metalinguistic Explanations) on their Written Production? . 138
  5.3 Research Question 3: How do Low-proficiency L2 Chinese Writers Engage with the Two Different Types of WCF (Direct vs. Indirect with Metalinguistic Explanations)? .................................................. 144
  5.4 Research Question 4: What are Teachers’ Perspectives about the Two Different Types of WCF (Direct vs. Indirect with Metalinguistic Explanations)? .............................................................. 146

CHAPTER VI: LIMITATIONS, CONCLUSION, AND FUTURE RESEARCH ..................... 148
  6.1 Limitations ........................................................................................................... 148
  6.2 Conclusion ........................................................................................................... 152
  6.3 Future Direction .................................................................................................. 154

APPENDICES ........................................................................................................... 158
Appendix A: Participant Pre-semester Learning Needs Questionnaire ........................................ 158

Appendix B: Online Task-based Curriculum Mandarin Chinese Express ........................................ 159

Appendix C: Writing Task 1 ........................................................................................................ 162

Appendix D: Writing Task 2 (Delayed Posttest) ........................................................................ 164

Appendix E: End-of-Course Interview and Survey ...................................................................... 166

REFERENCES ................................................................................................................................ 171
LIST OF TABLES

Table 1. Overview of Studies Comparing Focused VS. Unfocused Written Corrective Feedback .......... 19
Table 2. Overview of Studies Comparing Focused Direct, Indirect, and Indirect Written Corrective Feedback with Metalinguistic Explanations .......................................................... 25
Table 3. Overview of Studies Comparing Unfocused Direct, Indirect, and Indirect Written Corrective Feedback with Metalinguistic Explanations .......................................................... 32
Table 4. Examples of Three Levels of Depth of Processing (DoP) ........................................................................ 41
Table 5. Overview of Process Procedures ........................................................................................................ 43
Table 6. Overview of Studies Employing Language Related Episodes, Written Languaging, and Think-Aloud Protocols .................................................................................................... 54
Table 7. Student Participants Biographic Data .................................................................................................. 65
Table 8. Indirect WCF with Metalinguistic Explanation Sheet ........................................................................ 81
Table 9. Depth of Processing Coding Scheme .................................................................................................. 86
Table 10. Descriptive Statistics for Particle Le Accuracy Scores by Type of WCF ................................................. 92
Table 11. Summary of Between-Subject Analysis of Variance for Particle Le .................................................. 93
Table 12. Summary of Within-Subject ANOVA of Accuracy Scores ................................................................. 93
Table 13. Post-hoc Pairwise Comparison for Between Groups Particle Le Accuracy Scores on Revision of Writing Task 1 and New Writing Task 2 ........................................................................... 95
Table 14. Effect Size for Le Mean Accuracy ....................................................................................................... 95
Table 15. Post-hoc Pairwise Comparison for Within WCF Groups Particle Le Accuracy Scores on Writing Task 1, Revision of Writing Task 1 and New Writing Task 2 ......................................................... 96
Table 16. Descriptive Statistics for Classifier Accuracy Scores per Written Corrective Feedback Type .. 97
Table 17. Summary of Between-Subject Analysis of Variance for Classifiers ..................................................... 99
Table 18. Summary of Within-Subject Analysis of Variance for Classifier Accuracy Scores ................. 99
Table 19. Post-hoc Pairwise Comparison for Between Groups Classifier Accuracy Scores on Revision of Writing Task 1 and New Writing Task 2

Table 20. Effect Size Between Groups for Classifiers

Table 21. Post-hoc Pairwise Comparisons for Within WCF Groups for Classifier Scores

Table 22. Depth of Processing Results for Particle Le per WCF Type

Table 23. Examples of Direct WCF with Low Level of Depth of Processing on Particle Le

Table 24. Examples of Direct WCF with Medium-Level of Depth of Processing on Particle Le

Table 25. Examples of Direct WCF with High-Level of DoP on Particle Le

Table 26. Examples of Low-Level of Depth of Processing on Particle Le: Indirect WCF with Metalinguistic Explanations

Table 27. Examples of Medium-Level of Depth of Processing on Particle Le: Indirect WCF with Metalinguistic Explanations

Table 28. Examples of High-Level of Depth of Processing on Particle Le: Indirect WCF with Metalinguistic Explanations

Table 29. Percentage of Participants by Level of Depth of Processing for Classifiers

Table 30. Examples of Direct WCF with Low-Level of Depth of Processing on Classifiers

Table 31. Examples of Direct WCF with Medium-Level of Depth of Processing on Classifiers

Table 32. Examples of Direct WCF with High-Level of Depth of Processing on Classifiers

Table 33. Examples of Low-Level of Depth of Processing on Classifiers: Indirect WCF with Metalinguistic Explanations

Table 34. Examples of Medium-Level of Depth of Processing on Classifiers: Indirect WCF with Metalinguistic Explanations

Table 35. Examples of High-Level of Depth of Processing on Classifiers: Indirect WCF with Metalinguistic Explanations

Table 36. Overview of Detailed DoP Analysis Participant Background with Le Examples

Table 37. Overview of Detailed DoP Analysis Participant Background with Classifier Example
LIST OF FIGURES

Figure 1. Types of Written Corrective Feedback Overview ................................................................. 12
Figure 2. Feedback processing framework based on Leow’s (2015) Model of the L2 Learning Process in
Instructed Second Language Acquisition .......................................................................................... 39
Figure 3: Ellis’s (2010) WCF Engagement Relationship in Zheng and Yu’s (2018) study .................. 56
Figure 4. Online Course Material for Individual Study ........................................................................ 67
Figure 5. Adobe Connect Web Conferencing Software ..................................................................... 68
Figure 6. Chinese Task-base Curriculum .......................................................................................... 69
Figure 7. Procedure Timeline Overview ........................................................................................... 75
Figure 8. Writing Task 1 and New Writing Task 2 ............................................................................. 77
Figure 9. Participant’s error and unfocused direct WCF (in green) .................................................... 79
Figure 10. Participant’s error and unfocused indirect WCF (with codes and examples in sheet) ....... 80
Figure 11. Think-Aloud Protocol Training Material .......................................................................... 83
Figure 12. Depth of Processing Score Ranges (Zamora, 2017) ......................................................... 88
Figure 13: Formula for Target-Like Use Analysis of Morphemes from Pica (1983) ......................... 90
Figure 14. Boxplot for Particle Le Accuracy Score .......................................................................... 92
Figure 15. Particle Le Accuracy Score by Type of WCF ................................................................. 94
Figure 16. Boxplot for Classifier Accuracy Score ............................................................................. 98
Figure 17. Classifiers Accuracy Score by WCF Type ..................................................................... 100
Figure 18: Depth of Processing Results for Particle Le per WCF Type .............................................. 105
Figure 19. Depth of Processing Results for Classifiers per WCF Type ............................................ 111
Figure 20. Le Depth of Processing on the First Ten Errors ................................................................. 117
Figure 21. Classifier Depth of Processing on the First Ten Errors ..................................................... 120
CHAPTER I: INTRODUCTION

1.1 Statement of the Problem

Task-based language teaching (TBLT) is a popular approach to second language pedagogy that advocates for second language learning through the practice of authentic, real-life tasks (e.g., East, 2012; Long, 2016). Tasks in TBLT are generated from needs analyses of how students will apply their language skills outside the classroom and serve as the building blocks of a task-based curriculum (e.g., Long, 1985; 2015). A variety of prior work has demonstrated that task-based language teaching is effective in different language programs and language contexts, but there are still gaps in the empirical understanding of how best to implement task-based curriculums (e.g., González-Lloret & Nielson, 2015; Bryfonski & McKay, 2017). Recent research has examined task features that influence language learning and provided evidence for specific task designs and implementations, however, most task-based research has focused on oral tasks and task output practices (Plonsky & Kim, 2016). Written tasks have received less attention in the field of task-based language teaching theory and research (Gilabert, Manchón & Vasylets, 2016), despite documented writing needs from language students (e.g., Oskoz & Elola, 2014). Task-based language teaching research could benefit from an exploration of how writing and the processing of written corrective feedback (WCF) can influence second language (L2) learning in a task-based curriculum.

To examine writing within task-based language teaching, it is important to first look to the broader second language acquisition (SLA) and L2-writing research agendas. Writing has been shown to be an effective approach to language learning (Manchón, 2011c, 2018), especially where L2 writers can engage with feedback. Writing in an L2 is difficult and overall accuracy is relatively hard to achieve. Vocabulary, grammar, sentence order, and style all pose challenges
for students who need to meet voice, audience, and genre-based requirements that may be more abstract than in oral language (e.g., Hinkel, 2002; 2004; Silva, 1993). These challenges can make empirical measurement of improvements even more difficult to operationalize.

In order to improve writing accuracy and maximize language development, many studies in instructed second language acquisition have focused on the application of WCF in the classroom, leading to investigations of the potential effects of different types of feedback—for example, focused versus unfocused feedback; direct versus indirect feedback; or feedback with and without metalinguistic explanations. More specifically, focused WCF addresses only one target item in the whole composition while unfocused WCF reflects no focus on one or a few forms. Direct WCF provides the correct form of the error while indirect WCF highlights an error without any further information. Indirect WCF may also be accompanied by further metalinguistic information on the source of the error. More detailed discussion and examples of these feedback types are presented and discussed in the next section.

Many studies have demonstrated that WCF can improve L2 writing accuracy, complexity, and fluency (e.g., Bitchener, 2008; Ellis, Sheen, Murakami, & Takashima, 2008; Ferris, 2006; Russell Valezy, & Spada, 2006; Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause, & Anderson, 2010). However, a few researchers have argued that all kinds of WCF are ineffective (e.g., Kepner, 1991; Truscott, 1996, 2007; Truscott & Hsu, 2008). Overall, however, the literature remains unclear as to what types of feedback may be most useful in different contexts. Researchers have argued that focused WCF yields better results because L2 writers can only pay attention to one type of error at a time, while also saving teachers time and energy (e.g., Ellis et al., 2008; Sheen, Wright, & Moldawa, 2009). Other researchers have argued that direct WCF is more effective because L2 writers can easily juxtapose positive evidence against their
errors (e.g., Karimi, 2016; Benson & Dekeyser, 2018, Cerezo, Manchón, & Nicolás-Conesa, 2019). Prior research has also argued that indirect WCF is more effective because this type of feedback engages L2 writers in a problem-solving activity that encourages them to think more deeply (e.g., Lalande, 1982; Ferris, 1995; Bitchener & Knoch, 2008; Hartshorn et al., 2010).

Despite these prior investigations of WCF, it should be clear from the lack of general agreement in findings that there continue to be many unexplored areas ripe for investigation in this domain. Authentic classroom research is sparse, with laboratory findings dominating the literature. The challenges of the classroom are distinct, with anecdotal cultural bias among many teachers that adhere to unfocused WCF conventions that turn them into “error hunters” (Hairston, 1986, p. 122). It can be unclear for teachers how to approach the volume of errors their learners make or how to best target corrections (e.g., Storch & Wigglesworth, 2010; Hyland & Wong, 2013; Lee, 2018). Furthermore, most WCF studies have yet to investigate how participants processed the different types of feedback provided on their errors (Leow, 2019). Only a few studies have analyzed how participants noticed and reacted to the written feedback they received (Han & Hyland, 2015), and how they subsequently corrected the error and modified output (Suzuki, 2012, 2017; Bitchener, 2018; Manchón, 2018; Caras, 2019; Cerezo et al., 2019). Further research in this area could shed light on the effectiveness of different types of WCF, providing a more complete understanding of how L2 writers engage with different types of feedback, and address how feedback fits into the complexity of an authentic classroom environment.

In order to address some of these under-studied areas in TBLT and writing research, the current study examined the development of L2 Mandarin Chinese writing accuracy following WCF treatments in a task-based curriculum context. The study combines quantitative analysis of
accuracy improvements in writing on different targets, and builds on the use of process measures employed by previous studies (Caras, 2019; Han & Hyland, 2019) to provide more nuanced insights into how these L2 Chinese writers processed different types of WCF. The study applied think-aloud protocols, a type of concurrent data elicitation procedure that asks L2 learners to express their thoughts aloud in an effort to capture internal cognitive processes while they are interacting with the L2. The study triangulated qualitative data from think-aloud protocols and interviews, with quantitative analysis of accuracy scores to capture L2 Chinese writers’ writing accuracy developments and their processing of different types of WCF.
CHAPTER II: LITERATURE REVIEW

This chapter presents an overview of previous studies and theoretical frameworks that motivate the study. The first section covers the foundations of the task-based language teaching method and provides relevant context on the program that hosted the current study. The second section provides an overview of studies examining the effectiveness of different types of written corrective feedback in different contexts. The third section summarizes studies that have examined how written corrective feedback is processed by participants. The fourth section reviews studies addressing teacher perspectives and L2 writers’ engagement with written feedback. The chapter ends with a summary of the related literature and summarizes the research questions.

2.1 Task-Based Language Teaching

The task-based approach is a learner-centered, experiential approach to second language pedagogy that is grounded in the cognitive-interactionist theories of SLA (Long, 1985; Bygate, 2000). In the language classroom, task-based language teaching emphasizes the use of “tasks”—the “communicative uses to which learners will apply the L2 beyond the classroom, the things they will do in and through the L2” (Long, 2016, p. 6). Task-based language teaching aims to prepare learners to meet their current and future real-world communicative needs, including through writing. Learners’ needs are identified through the needs analysis process and practiced through pedagogic tasks that mimic and decompose the target tasks (Long & Norris, 2000; Long 2015; Ellis, 2003). Task-based language teaching is a very popular teaching approach in many language programs and language contexts (Bryfonski & McKay, 2017), including in the synchronous computer-mediated communication context examined in the current study (González-Lloret & Ortega, 2014). Task-based language teaching has also shown promise as an
innovative method for curriculum design, offering concepts and frameworks for developing various aspects of language curricula (Long & Crookes, 1992; Long & Norris 2000; Doughty & Long 2003; Chapelle, 2012; Mislevy & Yin, 2012). However, as uncovered in a recent meta-analysis on task-based learner production (Plonsky & Kim, 2016), there is limited linguistic diversity in the research base—the majority of task-based studies investigate English as the target language. As of January 2020, there are only four task-based studies that have examined Chinese as the target language (Lai, Zhao & Wang, 2011; Hill & Tschudi, 2011; Winke, 2014; Nielson, 2014).

Lai, Zhao and Wang (2011) investigated 36 novice L2 Chinese learners and four teachers’ reactions to the implementation of a task-based curriculum. Pre- and post- oral exam data demonstrated that the task-based participants performed significantly better than the control group. Surveys and interviews revealed both students and teachers reacted positively to the task-based learning and teaching experience. Lai, Zhao, and Wang (2011) showed the potential for tasked-based curriculum and speaking tasks in the online context, however, they did not report any data or procedure for needs analysis. It was unclear if this critical step was conducted, as it might have been deemed impractical given the course constraints. Needs analysis plays a foundational role in task selection and syllabus design, and the lack of a needs analysis limits the interpretability of TBLT as an approach in this study.

Hill and Tschudi (2011) conducted needs analysis and added task-based activities to a textbook-based, blended-context program. Twenty-one L2 intermediate Chinese learners participated in this study. Recordings of participants and teachers’ task conversations and survey results showed evidence that the TBLT activities supported L2 development by promoting feedback noticing and uptake. Hill and Tschudi (2011) improved upon previous studies by
incorporating needs analysis and designing related speaking tasks to facilitate participants’ oral development in a blended learning context. However, when they evaluated participants’ performance, Hill and Tschudi only used grammatical accuracy as the major measurement rather than a measure of task achievement or task performance. The inclusion of a control group and comparisons of task performance would have added valuable additional evidence regarding the effectiveness of the task-based activities.

The use of task-based assessments in task-based courses or programs has also been a topic of focused research (Chalhoub-Deville, 2001; Norris 2009; 2015), however, there are limited implementations available in prior literature. Winke (2014) explored task-based formative assessments to help learners to self-monitor their oral development in a 15-week online Chinese course. Seventeen advanced L2 Chinese learners participated in the study. Winke compared L2 Chinese participants’ rating and native speakers’ ratings on two oral tasks. She found that participants were realistic judges of their task performance, even when compared to experts, concluding that task-based assessments should be brought into the language classroom.

Nielson (2014) evaluated an intermediate online task-based Chinese course through task-based assessments over one year. Thirty-five participants joined the task-based teaching program and 15 participants were in the control group. Both groups completed pre- and posttests of Chinese reading and speaking. The task-based teaching group improved significantly on both speaking and reading in the post test. Because only 9 participants finished pre and post-test, the sample size was not big enough to conduct any statistical testing to compare the task-based group and the control group. A survey of the task-based course results indicated both participants and teachers were very satisfied with the course and noticed improvements.
These four highlighted case studies demonstrate that task-based language teaching was well accepted by both students and teachers in L2 Chinese contexts, but the focus of these four case studies was limited to oral tasks and did not investigate written tasks. A more holistic examination of task-based language teaching should also include writing tasks with WCF and examine the effects for target language development. Bruton (2002) pointed out that task-based language teaching tends to overemphasize speaking tasks and spoken production since task-based approaches arose from communicative language teaching. Bruton argued that beginners cannot learn much from working on speaking tasks, especially tasks that include new language or linguistic knowledge. Wigglesworth (2001) and Robinson (2011) also expressed concern with TBLT’s emphasis on learners’ oral language development. In a recent meta-analysis, Plonsky and Kim (2016) examined 85 studies (2006-2015) on task-based L2 learner production. Each study was coded for target features (contextual, demographic, and methodological features). Plonsky and Kim found there was a strong preference toward interactional features of oral tasks. Little research interest was shown toward other task modalities and feedback types. Therefore, they recommended future research pay more attention to other task types, task quality and variety of learner demographics.

Besides oral tasks, researchers have argued for designers and practitioners to develop tasks that encourage written production since writing tasks offer greater visibility of task performance and corrective feedback. (Ellis, 2009; Bygate, Skehan, & Swain, 2013). For example, Macaro (2014) drew on cognitive theories of writing research and conducted a case study that explored how two L2 French learners used different writing strategies during task performance. Macaro (2014) argued that task-based language teaching can be expanded from a focus on task manipulation and performance to involvement by examining complex cognitive
writing activity. Byrnes and Manchón (2014) suggested that task-based language teaching and L2 writing can inform and enrich each other and argue that little attention has been paid to investigating tasks as research and learning tools within the L2 writing domain. In the future, they suggest task-based language teaching research studies can be thematically situated in the learner-internal perspectives, learning orientation, textual meaning-making orientation, and curricular orientation.

Finally, Gilabert, Manchón and Vasylets (2016) argued that the role of task mode was under-researched and called for a more integrative and mode sensitive task-based language teaching research agenda. They explained that oral and written tasks represented different language learning environments and learning processes, while few studies attempted to explore the potential of written tasks and WCF. Written tasks are slow paced and visible, as opposed to oral tasks, which are ephemeral. These features allow learners to take time to pay attention to input and facilitate written feedback processing. They also recommended applying Leow’s (2015) model of the learning process in ISLA to give structure to task performance-related processes since Leow’s model provides all the processes and products that can be generated by tasks. (Refer to Section 2.3 for a more detailed discussion of Leow’s model.) Therefore, more attention should be paid to writing and WCF in task-based language teaching research and theory. Investigating writing tasks and WCF can help practitioners identify task features that would promote language processing for comprehension and learning, hence providing empirical evidence for task design and teachers’ practices.

The task-based language teaching literature emphasizes using real-world authentic tasks to engage learners in meaningful interactions, but the majority of attention has been paid only to spoken tasks. Scant attention has been paid to investigating the written mode of tasks. The
written mode provides ample opportunities for researchers and teachers to design written tasks and examine their effect on written task productions.

2.2 Effectiveness of Different Written Corrective Feedback Types

Written corrective feedback is form-focused feedback, a type of external manipulation of an L2 learners’ written product to draw his or her attention to errors (e.g., Bitchener & Storch, 2016; Ellis, 2009; Shintani, Ellis, & Suzuki, 2014). There are five major types of WCF that teachers can provide to students on their written text: focused, unfocused, direct and indirect with or without metalinguistic explanations. Researchers have investigated various types of WCF and reported conflicting results on its effectiveness over the last three decades (Benson & DeKeyser, 2018). The section below provides exemplars of these three different types of WCF from the current study. A student who wanted to express “I bought a bottle of water at the supermarket.” produced the sentence in Mandarin Chinese below. Their production in Chinese characters is accompanied by the phonetic romanization using the pinyin system, with a literal word-for-word translation, as well as an English translation below. Words in red are errors, while words in green are examples of type of WCF provided by the teacher.

Example 1

Chinese: 我买一个水在超市。

Pinyin romanization: Wǒ mǎi yī gè shuǐ zài chāoshì.

Literal translation: I / buy / one / classifier / in / supermarket /.

Translation: I buy a water in the supermarket.

There are three errors in this sentence. First, the particle le should be added after the verb to indicate that the action of buying is completed. Second, ge is the wrong classifier; the correct
classifier is *ping* (bottle). Third, the word order of the sentence is incorrect. Location of the action should be placed in front of the main action verb in Mandarin Chinese. Therefore, *zài chāoshi* (in the supermarket) should precede *mǎi* (to buy).

When a teacher provides WCF on only one target item in the whole composition, this type of WCF is called focused WCF. An example of this type of focused correction is shown below with a focus on classifiers as the target grammatical item.

**Example 2**

Chinese: 我买一个瓶水在超市。

Pinyin romanization: *Wǒ mǎi yī gè píng shuǐ zài chāoshì.*

Literal translation: I / buy / one / classifier / classifier bottle / water / in / supermarket /

Translation: I buy a bottle of water in the supermarket.

In the example above, the teacher focused on the classifier error and only gave WCF on *ping* (bottle). The verb aspect and word order errors are ignored, that is, no WCF related to verb aspect and word order was provided. This type of feedback is considered focused WCF since it only focuses on one or a very limited range of errors.

In comparison, when a teacher provides WCF on a wide range or all types of errors, this practice is considered unfocused WCF. In unfocused WCF, all the errors are marked and corrected. An application of unfocused feedback on the same error is provided below.
Example 3

Chinese: 我在超市买了一个瓶水在超市。

Pinyin romanization: Wǒ zài chāoshì mǎi le yī gè píng shuǐ zài chāoshì.


Translation: I bought a bottle of water in the supermarket.

Besides the classifier error, the teacher also corrected the verb aspect, changing mǎi to mǎi le to express the past-tense “bought” in Chinese; they also corrected the word order by moving zài chāoshì (in the supermarket) to the front of the action verb.

In studies of focused WCF, L2 writers only receive feedback on a few specific forms. Studies have examined English articles (Bitchener, 2008; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007), hypothetical conditionals (Shintani, Ellis, & Suzuki, 2014), prepositions (Ellis et al., 2008), and the past simple tense (Bitchener, Young, & Cameron, 2005). In contrast, studies of unfocused WCF address a range of error types in a piece of writing, which can include punctuation, missing words, verb forms, speech order, spelling, and other categories (e.g., Lalande, 1982; Robb, Ross, & Shortreed, 1986; Chandler, 2003; Mubarak, 2013).

![Figure 1. Types of Written Corrective Feedback Overview](image-url)
Another line of WCF research explores the effects of different types of WCF: direct and indirect with or without metalinguistic explanations. These can be considered on an axis of the directness of the feedback, starting from clear and concrete direct feedback on the left (see Figure 1). Direct feedback consumes the most teacher time as they replace each error with a correct form. Indirect feedback without metalinguistic explanations sits on the opposite end as the least direct, only marking an error and not providing additional direction with regard to the source of the error. This is the least intensive for the teacher, who does not need to produce any additional markings for errors. Indirect feedback with metalinguistic explanations (indirect+) sits in the middle of directness and time intensity. Three examples from the current study are provided below to illustrate these feedback type concepts. The teacher provided each of the three types of WCF on participants’ written production of L2 Mandarin Chinese. Mandarin characters are represented in pinyin romanization.

Direct feedback is provided by crossing out the error and inserting explicit corrections. In Example 4, a learner made a lexical error on the Chinese classifier bottle. In direct WCF, each error is marked and corrected by the teacher.

**Example 4**

Chinese: 我买了一个瓶子。

Pinyin romanization: Wǒ mǎi le yī gè píng shuǐ.

Literal translation: I / buy / Particle le / one classifier (general) / classifier (bottle) / water

Translation: I bought a piece bottle of water.
Indirect WCF with metalinguistic explanations does not give corrections, but points out the error and provides metalinguistic explanations on how to correct the error. This type of explanation may be provided on a separate sheet with the meanings of the encoded indirect feedback, as shown below on the line for metalinguistic feedback.

In Example 5, the L2 writer wanted to express “I went to the school” and the action of “to go” is already finished. They forgot to use *le* in their sentence to mark past tense, and without it, *Wǒ qù xuéxiào* means “I go to school.” The verb aspect error was marked with indirect WCF. This type of feedback appeared with the metalinguistic code “P-le1” alongside the metalinguistic explanation for the L2 writer to interpret. Direct correction was not provided. The L2 writer must figure out how to correct this error based upon the metalinguistic code and explanation.

**Example 5**

Chinese: 我去 P-le1 学校。

Pinyin romanization: *Wǒ qù P-le1 xuéxiào.*

Literal translation: I / go / P-le1 / school.

Translation: I go to the school and verb “go” has the error code “P-le1”.

Metalinguistic Feedback: P-le1 means particle *le* usage 1. *Particle le is used to express action completion, such as Wǒ chī le fàn (I ate dinner).*

Indirect WCF without metalinguistic explanations only points out the error without correction, providing a marking, but no correct form or explanations for the L2 writer. An example of this is shown below where the classifier *ge* was underlined in green to show that
there was an error there, but no additional information was provided as to whether it was an error in word choice, placement, or other grammatical form.

Example 6

Chinese: 我买了一个水。

Pinyin romanization: Wǒ mǎi le yī gè shuǐ.

Literal translation: I / buy / particle le / one classifier (general) / water.

Translation: I bought a piece of water.

Direct WCF, indirect WCF with metalinguistic explanations, and indirect WCF all have strengths and weaknesses. Teachers have frequently employed all of them in authentic classroom contexts based on their perceptions and personal preferences. Researchers from both the instructed second language acquisition and writing fields have investigated these three types of WCF to explore which types of feedback may be better in certain situations, but the findings are still inconclusive. The next section summarizes studies that compared focused direct, indirect + and indirect WCF.

2.2.1 Focused vs. Unfocused Written Corrective Feedback

A variety of prior studies have empirically investigated the effects of focused versus unfocused WCF. Ellis et al. (2008) examined 49 L1 Japanese participants’ usage of the English indefinite and definite articles in their compositions. There were two feedback treatment groups and a control group. The focused WCF group received correction on only English article errors on three written drafts and the unfocused written feedback group received correction on multiple error types, including English past tense, articles, prepositions, and vocabulary. Besides these
three writing tasks, they also added a grammar correction test that was adopted from Sheen (2006). They found both feedback treatment groups improved on both the grammar correction test and writing tasks (posttest). Both treatment groups also outperformed the control group. The difference between the two treatment groups was not statistically significant. Surprisingly, Ellis et al. (2008) argued that focused WCF would be more effective based on the mean accuracy score improvement from posttest to delayed posttest. Unfortunately, Ellis et al. (2008) did not report within groups comparison statistics or any effect size. Without any statistical analysis or qualitative data, their suggestion to use focused WCF should be considered with caution.

Sheen, Wright, and Moldawa (2009) also investigated the effectiveness of focused and unfocused WCF on the usage of English articles, as measured on a writing task and grammatical tests. Eighty EFL participants were assigned to four groups: a focused WCF group, an unfocused WCF group, a writing practice group (which only wrote compositions without any WCF), and a control group that only completed an articles grammar test. Sheen et al. (2009) found that both treatment groups significantly outperformed the control group. The focused WCF group was significantly better than the unfocused group. Surprisingly, the focused WCF group did not only achieve the highest accuracy score gain on English articles, but also on the other grammatical structures, such as regular past tense, irregular past tense, and prepositions. Sheen et al. (2009) explained that the focused group participants had paid attention to the direct corrections while the unfocused group participants were unable to process the wide range of grammatical errors effectively. There were four weaknesses in Sheen et al.’s (2009) study. First, there were methodological issues with the writing task. The teacher read a story in English, and the participants wrote the story based on their listening skills. It was therefore possible that a participant did not write correctly because of a gap in listening comprehension or because they
did not remember the story after the teacher’s reading. Second, from the study’s descriptive statistics, the focused group had the highest pretest score, so it is possible that the groups were not equivalent at pretesting. No post-hoc analyses were reported between different group pairings to ensure that there was no statistical difference on the baseline pretest. Third, the researchers did not test the participants’ accuracy on a new piece of writing to demonstrate that they acquired the knowledge from the written corrective feedback. Lastly, even though Sheen et al. argued that different groups processed WCF differently, they did not employ any process measures and did not include any concurrent data to show how participants processed WCF.

Shepherd, Daily O’Meara, and Snyder (2016) examined 279 English as a second language participants’ writing accuracy improvement over a semester. Their participants chose one of three feedback treatment groups: extensive unfocused direct feedback, focused direct feedback, and minimal unfocused direct feedback. They found no significant improvement for all three groups. The research was situated in an authentic curriculum; however, there were many methodological and reporting issues in this study. Shepherd et al. (2016) did not explain the writing task, the accuracy measurements, or the examples of the targeted “grammatical errors.” In addition, the participants who received extensive feedback were asked to have a discussion with the teachers. Participants who received focused feedback had the option of having a discussion with the teachers. The minimal group did not have the discussions. The discussions with the teachers lowered the level of internal validity of the research design because it was unknown what caused participants’ writing accuracy change—the treatment or some element of the discussion.

As summarized in Table 1, researchers have argued that focused WCF yields better results than unfocused WCF for intermediate L2 writers (e.g., Ellis et al., 2008; Sheen et al.,
however, this difference may arise from a greater ability to target and measure how participants improved given that this research focuses precisely on the accuracy change of one or two target items—for example, the correct usage percentage of classifiers in participant’s drafts, allowing for clear statistical comparisons (e.g., Bitchener & Knoch, 2009). On the other hand, studies of unfocused WCF have employed a variety of general measures of accuracy that may not be internally comparable among linguistic targets—for example, an error-free T-unit ratio as in Hartshorn et al. (2010), error per 100 words, or error-to-words ratio (Van Beuningen et al., 2012). It is difficult to track accuracy improvements though these kinds of holistic error measurements. For example, a participant could make 20 errors in 100 words in the first draft, and these 20 errors could occur in error type A, B, and C. In a second draft after unfocused WCF, it is possible the student would still make 20 errors in 100 words. However, the student may improve in accuracy on type A, B, or C, but make new errors of type E, F, and G. From this hypothetical example, even though the student improved writing accuracy on the old target-type errors, their total writing accuracy remained the same because of the holistic measurement method.

From a foreign language curriculum perspective, the purpose of writing tasks is to provide opportunities to practice what has been covered in a lesson or unit, often including a broad range of grammar and vocabulary (Manchón, 2011). Feedback cannot be limited to only one or two linguistic targets in the real-world language classroom, so unfocused WCF treatments may better replicate what many practitioners do and acknowledge the complex features of writing, such as structure and content (Lee, 2018). As Van Beuningen, De Jong, and Kuiken (2012) argue, focused WCF studies may oversimplify the writing task as a grammar exercise and may cause learners to intentionally focus on the linguistic targets.
Table 1. Overview of Studies Comparing Focused VS. Unfocused Written Corrective Feedback

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target measurements</th>
<th>WCF types</th>
<th>Process measures</th>
<th>Accuracy Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ellis et al. (2008)</td>
<td>49 ESL (intermediate)</td>
<td>English referential Articles</td>
<td>-Focused direct</td>
<td>None</td>
<td>Both feedback treatment groups outperformed the control on the posttests and the grammar tests. The difference between the two treatment groups was not statistically significant.</td>
</tr>
<tr>
<td>Sheen et al. (2009)</td>
<td>80 EFL (intermediate)</td>
<td>English referential Articles</td>
<td>-Focused direct</td>
<td>None</td>
<td>All the treatment groups outperformed the control group on overall accuracy. The focused group outperformed others in article accuracy.</td>
</tr>
<tr>
<td>Shepherd et al. (2016)</td>
<td>279 ESL (intermediate)</td>
<td>The top 1-3 repeated errors were custom for each participant in the focused group.</td>
<td>-Focused direct</td>
<td>None</td>
<td>No improvements.</td>
</tr>
</tbody>
</table>
2.2.2 Focused Written Corrective Feedback

In a study of focused WCF, Sheen (2007) examined 91 intermediate English-as-a-second-language (ESL) participants in two focused WCF conditions (direct, direct with metalinguistic explanation, and a control group) with a pre-test, treatment, post-test, and delayed posttest research design. The target structures were the definite and indefinite English articles “the” and “a.” The teachers only provided WCF on errors in those target articles and ignored errors in other areas. The control group did not perform any writing tasks, only completing the grammar tests. Sheen found that both feedback groups performed better than the control group. From pre-test to post-test, the group differences were not significant. In the delayed post-test, only the group receiving direct with metalinguistic explanations performed significantly better than the control group. The post-hoc comparisons showed significant differences among the three groups in the delayed post-test. The direct with metalinguistic explanations feedback group outperformed the direct feedback group on both posttest and the delayed posttests. From these results, Sheen argued that WCF is beneficial to English article acquisition. Sheen also claimed that the feedback with metalinguistic explanation was better than direct correction only, and this could be explained using Schmidt’s (2001) role of awareness in L2 acquisition. According to Schmidt, there are two levels of awareness: at the level of noticing where learners only pay attention with a low level of awareness, and the level of understanding where the conscious awareness of rules arises. Direct feedback without metalinguistic explanation may lead students to process at the level of noticing. However, direct feedback with metalinguistic explanations may push students to reach the level of understanding since understanding the grammatic rules for English articles has been shown to promote later acquisition (e.g., Ellis et al., 2006).
Bitchener, East, and Cartner (2010) sought to address whether indirect feedback with metalinguistic explanations would be more beneficial when compared with direct WCF. They investigated 20 advanced ESL learners’ acquisition of English nouns and subject-verb agreement. The study results showed that both direct and indirect-with-metalinguistic-explanations groups improved significantly in the writing accuracy on noun number and there was no difference in performance between the two types of feedback. Only indirect feedback with metalinguistic explanations improved the participants’ accuracy in subject-verb agreement. In Bitchener et al. (2010), authors only reported the results without any analytical discussion and the effect sizes were not reported. This study had a low n size and the control group was absent. Without this detailed information, it is difficult to understand why the indirect-with-metalinguistic-explanations group outperformed the direct WCF group.

In another study comparing types of WCF, Shintani, Ellis, and Suzuki (2014) compared the effects of direct feedback and indirect feedback with metalinguistic explanations on 140 intermediate EFL participants’ acquisition of two grammatical items, indefinite articles, and the hypothetical conditional. Both direct WCF and indirect WCF with metalinguistic explanation treatment groups were split into subgroups: those who completed a revision of their writing and those who did not. Treatment groups without revision only looked at the feedback for five minutes, and treatment groups with revision rewrote the essay with the accompanying feedback treatments. Shintani et al. (2014) found that all the treatment groups improved on the hypothetical conditional, but not on the indefinite article. Only direct WCF with revision had a lasting effect on hypothetical conditional. They argued that linguistic structures have differences in saliency and complexity. Students paid attention to the hypothetical conditional because it is more salient and complex than “a” vs “an.” They also argued that the direct WCF with revision
would promote students’ attention to the more salient target. They concluded that direct WCF with revision was the most effective feedback in the study. However, this result should be viewed with caution due to the methodology associated with the delivery of the feedback in the indirect metalinguistic feedback group. The indirect-with-metalinguistic-explanations group did not receive any corrective feedback on the students’ first draft. This group only received metalinguistic explanations on the two target items and corrected their errors on their own. This unexpected format meant that students did not know where they made errors in their drafts, potentially making it quite confusing to figure out their errors on their own. This particular treatment is also inconsistent with the provision of indirect feedback in other studies and may make the results less generalizable.

Shintani and Ellis (2015) extended Shintani et al. (2014) and examined the role that learners’ language analytic ability played in the effects of the feedback types. They studied improvement in accuracy scores in a post-treatment writing task, and found higher language analytic ability was beneficial to groups receiving direct WCF and to groups receiving indirect WCF with metalinguistic explanations. They argued that students’ language analytic ability determines their ability to deduce the application of the pattern from the indirect metalinguistic explanation. Language analytic ability may also affect the processing of direct feedback by determining how a student can inductively construct the underlying grammatical rules.

Karimi (2016) not only examined direct and indirect WCF, but also added a group that received direct feedback first, followed by indirect WCF. He focused on 94 intermediate EFL students’ accuracy on English articles and simple past tense (regular and irregular). All participants completed a pre- and post-test (revision). The results showed all of the three treatment groups improved accuracy on the two target items, and differences in the three
treatment groups was significant with a large effect size of Cohen’s $d = 0.556$. In the post-hoc analysis, the direct group significantly outperformed the indirect group, and the group that received direct and indirect feedback. The difference between indirect and the direct following indirect group was not statistically significant. Karimi argued that direct WCF was most effective because participants not only received the indications of their errors, but also the correct forms. Even though the direct group had the highest accuracy rate in the study, this piece of evidence was not enough to prove the participants understood the feedback and took actions to correct their errors. It was possible that participants simply copied teachers’ feedback in their revisions. Information on how participants reflected on their errors and a delayed post-test could have provided greater insight into differing learner performances.

In an additional study, Benson and DeKeyser (2018) explored 151 English as a second language learners’ writing accuracy on simple past tense and present perfect tense under two WCF conditions: direct and indirect with metalinguistic explanations. They found both feedback groups performed better than a control group. However, while the direct WCF group evidenced a lasting effect on the simple past tense, the indirect feedback group with metalinguistic explanations did not. They suggested learners with higher language analytic ability would benefit more from direct WCF.

Table 2 provides an overview of studies comparing focused direct, indirect, and indirect-with-metalinguistic-explanations WCF. In summary, the body of research on focused WCF types demonstrates that both direct WCF and indirect WCF with metalinguistic explanations can be beneficial to intermediate English language learners. Direct WCF produced sustained accuracy gains on some linguistic structures (e.g., Shintani et al., 2014) while others (Bitchener & Knoch,
2010) found that indirect feedback with metalinguistic explanations led to greater accuracy than in the direct group.

In addition, focused WCF studies have investigated teachers’ feedback on a few specific forms, such as English articles (Sheen, 2007; Karimi, 2016) and past tense (Bitchener et al., 2010). Many studies have looked at the role of different WCF types alongside error treatability (e.g., Ferris, 1999; Sheen, 2007; Shintani et al., 2014). Ferris (1999), in a response paper to Truscott (1996), explained that the type of error might play a role in WCF’s effectiveness. If the type of error occurs in a patterned, rule-governed way, it is treatable and easier for participants to understand the rule and take actions upon their errors. On the other hand, if the type of error appears in a set of rules and irregular forms, participants can have a hard time understanding feedback and the correct forms. Following Ferris’s claim, Shintani et al. (2014) argued saliency and complexity of the grammatical structure are critical factors that should be taken into the consideration of WCF’s effectiveness. However, it is also critical to investigate how participants notice and process error feedback, especially feedback that varies in terms of saliency and complexity. Furthermore, prior research on focused WCF can be difficult to generalize to the authentic classroom settings where teachers provide feedback on a range of error types.
<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target measurements</th>
<th>WCF Types</th>
<th>Process measures</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sheen (2007)</td>
<td>91 ESL (intermediate)</td>
<td>English articles (a, an, the)</td>
<td>Direct - Direct + Control</td>
<td>None</td>
<td>Direct with metalinguistic explanations outperformed direct and control group on all three tests.</td>
</tr>
<tr>
<td>Bitchener, East, &amp; Cartner (2010)</td>
<td>20 ESL (advanced)</td>
<td>English singular/plural nouns, subject-verb agreement</td>
<td>Direct - Indirect +</td>
<td>None</td>
<td>Both groups improved in singular/plural nouns.</td>
</tr>
<tr>
<td>Shintani, Ellis, &amp; Suzuki (2014)</td>
<td>140 EFL (intermediate)</td>
<td>Hypothetical conditional (if), indefinite articles (an/a)</td>
<td>Direct - Indirect + Direct with revision</td>
<td>None</td>
<td>All treatment groups were better than control.</td>
</tr>
<tr>
<td>Shintani &amp; Ellis (2015)</td>
<td>118 EFL (intermediate)</td>
<td>Hypothetical conditional (if), indefinite articles (an/a)</td>
<td>Direct - Indirect + Direct with revision</td>
<td>None</td>
<td>An extension on Shintani et al. (2014) and results were the same.</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Target measurements</td>
<td>WCF Types</td>
<td>Process measures</td>
<td>Accuracy results</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------</td>
<td>--------------------------------------</td>
<td>----------------------------</td>
<td>------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Karimi (2016)</td>
<td>94 EFL (intermediate)</td>
<td>Articles, simple past tense</td>
<td>Direct, Indirect, Direct then indirect, Control</td>
<td>None</td>
<td>All treatment groups were better than control. There was a difference in effect between three types of feedback.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Direct was the best, followed by “direct then indirect”, the last one was “indirect”. Direct was better than indirect for simple past tense.</td>
</tr>
<tr>
<td>Benson &amp; Dekeyser (2018)</td>
<td>151 ESL (Intermediate and Advanced)</td>
<td>Simple past tense, the present perfect tense.</td>
<td>Direct, Indirect +, Control</td>
<td>None</td>
<td>For the present perfect tense, both improved on a new piece of writing, but neither group maintained this improvement long-term.</td>
</tr>
</tbody>
</table>

“Indirect” means indirect WCF without metalinguistic explanations. “Indirect +” means indirect WCF with metalinguistic explanations.
2.2.3 Unfocused Written Corrective Feedback

Pivoting to unfocused WCF, many studies have provided feedback on many or all errors in the participants’ compositions. Lalande (1982) was the first study to compare indirect WCF with metalinguistic explanations and direct WCF in a foreign language classroom context. Since direct WCF was the traditional classroom practice, Lalande described the direct group as the control group, and the indirect-with-metalinguistic-explanations group as the experimental group. Lalande tracked 60 L2 intermediate German students’ writing accuracy on 15 grammatical and lexical targets (such as subject verb agreement, word order, prepositions, negators, and nouns). The results indicated that the indirect-with-metalinguistic-explanations group significantly outperformed the direct feedback group on 11 grammatical targets, but not on lexical types of errors. Based on these results, Lalande suggested that L2 language teachers should consider using indirect WCF.

Semke (1984) assumed that unfocused WCF would increase students’ learning anxiety since the students’ written drafts were usually marked with a teacher’s red pen. In order to reduce learning anxiety, Semke added “comment only” and “peer correction” as other written feedback types in comparison with direct WCF. The “comment only” group only received comments on the content of the essay, and the “peer correction” group received written corrective feedback from other student participants. The study results showed no significant difference in participants’ writing accuracy, fluency, or complexity among these different WCF groups. Surprisingly, in the WCF attitude survey, students preferred the comments only group and dispreferred peer correction. Based on these results, Semke suggested that teachers do not need to use too much energy to correct L2 students’ written production. Instead, teachers should enjoy reading students’ compositions as mutual sharing of information and provide comments on
the content. If the teacher noticed consistent errors, then the errors should be addressed in the drill sessions, according to Semke. There was an issue in the “comment only” treatment group—this group wrote twice as much as other groups since participants in the “comment only” group were incentivized by a different grading scale. Whether the performance of the “comment only” group was superior to other treatment groups remains an unanswered question.

Robb, Ross, and Shortreed (1986) investigated saliency of different types of WCF. They examined all error types from 134 novice English as a second language students’ in four WCF groups: direct without metalinguistic explanation, indirect, indirect with metalinguistic explanation, and indirect with only error numbers on each line. The results showed that all the WCF groups improved their writing accuracy by having more error-free T units in their compositions. There were no significant differences among types of WCF. The authors suggested that a less time-consuming method to guide students’ attention to their errors, such as indirect WCF without metalinguistic feedback, might be better.

Chandler (2003) explored four types of WCF conditions: direct, indirect without metalinguistic explanations, indirect with metalinguistic explanation, and metalinguistic explanation on 36 English as a second language students’ writing accuracy, fluency, and complexity. Chandler found that all the treatment groups improved on writing accuracy. Direct WCF and indirect WCF without metalinguistic explanations significantly outperformed the indirect with metalinguistic explanation and metalinguistic explanation-only methods. However, Chandler’s (2003) research design was very different from other studies. In order to ensure a larger n-size for the study’s analyses, all 36 participants received each of the four WCF treatments in different orders. This design did not control for the potential time effects for each step in the sequence combined with interactions between the feedback types that could have
influenced the outcomes. Randomization of multiple treatments actually created dissimilar data sets that ideally would not be combined.

Ferris (2006) examined 92 intermediate ESL students’ writing accuracy on 15 grammatical and lexical error types. The participants received all four types of WCF: direct, indirect without metalinguistic explanation, indirect with metalinguistic explanation and comments. She found WCF significantly improved students’ writing accuracy. Teachers preferred to use direct and indirect WCF with metalinguistic explanation methods. Ferris was also very interested in error treatability, finding errors that were easier to correct were often marked indirectly, and untreatable errors (such as sentence order) were overwhelmingly corrected using the direct WCF.

Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause, and Anderson (2010) investigated the effectiveness of unfocused direct WCF and indirect WCF with metalinguistic explanation in an English as a foreign language writing course with 47 advanced L2 English learners. They assigned participants to two groups and provided them with one type of unfocused WCF. They compared rhetorical competence and writing CAF (complexity, accuracy and fluency) measures, finding that the group receiving indirect feedback with metalinguistic explanations outperformed the direct WCF group on accuracy, but with a “slight unfavorable effect on writing complexity and fluency.”

Caras (2019) examined 51 L2 Spanish students’ writing accuracy on ser versus estar and the preterit versus imperfect tense. There were three WCF treatment groups—direct, indirect without metalinguistic explanations, and indirect with metalinguistic explanations—and a control group. The most important difference between Caras’s study and other WCF studies lies in the use of think-aloud protocols to provide evidence of students’ cognitive processing of different
types of WCF. For ser versus estar, the direct group outperformed the other three groups on the revision of the first writing task; however, these significant differences did not last on a new task after two weeks. In comparing preterit versus imperfect tense, the type of WCF did not have a differential effect on accuracy over time. The think aloud data revealed that the majority of the direct and indirect-without-metalinguistic-explanation groups processed target linguistic items at a low level of processing while the indirect-with-metalinguistic-explanations group processed the targets at medium and high levels.

Cerezo et al. (2019) examined 46 undergraduates in two WCF conditions (direct and indirect only with metalinguistic codes) and a control group. Both treatment groups performed better than the control group, but there was not a statistically significant difference in the amount of successful corrections or errors left uncorrected in the revised texts. Cerezo et al. (2019) used written languaging to gather data on how students processed direct WCF and indirect WCF with metalinguistic codes. Written languaging is “the process of making meaning and shaping knowledge and experience through language” (Swain, 2006, P.98). The study found that the direct group had greater depth of processing, reaching to the level of understanding and explaining errors. In comparison, the indirect group only processed at a low level. By finding that WCF is not automatically processed fully and accurately, the study highlights the importance of further research into how WCF is perceived and processed by learners.

Table 3 presents an overview of these studies that examined unfocused WCF. Overall, studies comparing different types of unfocused WCF report mixed results. Many studies report no differences between these two WCF treatments (Caras, 2019; Cerezo et al., 2019; Ferris, 2006; Mubarak, 2013; Robb et al., 1986; Semke, 1984), while some researchers view indirect with metalinguistic explanations as more effective (Lalande, 1982; Hartshorn et al., 2010). These
studies suggest that indirect WCF may induce the learner to cognitively “engage in guided learning and problem solving” activities (Lalande, 1982) that foster long-term acquisition (Bitchener & Knoch, 2008). In other words, indirect WCF requires the learner to draw on their existing knowledge of the target linguistic form, incorporate new feedback information in working memory, and form a hypothesis about the accurate L2 output. However, whether L2 writers do indeed process such feedback, as assumed, has yet to be empirically investigated (Leow, in press).

This line of unfocused WCF research has some limitations in their collective research design, especially on accuracy measurement and WCF treatment. It is difficult to track participants’ accuracy improvement by using error-free T units, errors in 100 words, or all errors because they cannot ensure consistent measurement of error types. For example, it is possible that participants improved on some types of errors and made new types of errors in the next draft. In terms of WCF treatment, there is still only a limited amount of research in this area. Studies in this line adopted different WCF treatments making it difficult to compare and synthesize their results into a conclusion about which treatment may be better. Caras (2019) had two indirect WCF groups: indirect without metalinguistic explanations, and indirect with metalinguistic explanations. In Cerezo et al.’s (2019) study, they provided indirect WCF with metalinguistic codes to the participants. In Hartshorn et al.’s (2010) study, the researchers also included teacher discussions to both feedback treatment groups. With limited research and varying methods, there is still ample opportunity to explore the effects of different WCF approaches.
Table 3. Overview of Studies Comparing Unfocused Direct, Indirect, and Indirect Written Corrective Feedback with Metalinguistic Explanations

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target measurements</th>
<th>WCF Types</th>
<th>Process measures</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lalande (1982)</td>
<td>60 L2 German (Intermediate)</td>
<td>15 grammatic and lexical types</td>
<td>-Direct -Indirect +</td>
<td>None</td>
<td>“Indirect +” was better than direct group.</td>
</tr>
<tr>
<td>Semke (1984)</td>
<td>141 L2 German (Intermediate)</td>
<td>All errors</td>
<td>-Comments -Direct -Direct with comments -Peer correction</td>
<td>None</td>
<td>No group improved. There was not a difference in effect between four types of feedback. Participants like the comment- method best.</td>
</tr>
<tr>
<td>Robb et al. (1986)</td>
<td>134 ESL (Novice)</td>
<td>Lexicon, syntax, measured by error free T-units</td>
<td>-Direct -Indirect + -Indirect – -Indirect – (error numbers)</td>
<td>None</td>
<td>All treatment groups improved. No statistically significant differences among groups.</td>
</tr>
<tr>
<td>Chandler (2003)</td>
<td>36 ESL (Intermediate)</td>
<td>All errors types per 100 words</td>
<td>-Direct -Indirect + -Indirect – -Metalinguistic explanations only</td>
<td>None</td>
<td>All groups improved. Direct and indirect- were better than indirect + and metalinguistic group.</td>
</tr>
<tr>
<td>Ferris (2006)</td>
<td>92 ESL (intermediate)</td>
<td>15 types in verbs, noun endings, articles, word choice, and sentence structures</td>
<td>-Direct -Indirect + -Indirect -</td>
<td>None</td>
<td>Written corrective feedback improved students’ writing accuracy on the 15 error types. Direct and indirect + were more effective.</td>
</tr>
<tr>
<td>Study</td>
<td>Participants</td>
<td>Target measurements</td>
<td>WCF Types</td>
<td>Process measures</td>
<td>Accuracy results</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Hartshorn et al. (2010)</td>
<td>47 ESL (Advance)</td>
<td>All errors</td>
<td>-Direct -Indirect +</td>
<td>None</td>
<td>Indirect + was better than direct.</td>
</tr>
<tr>
<td>Mubarak (2013)</td>
<td>46 ESL (NA)</td>
<td>All errors</td>
<td>-Direct -Indirect + -Control</td>
<td>None</td>
<td>No improvement. No statistically significant differences among groups.</td>
</tr>
<tr>
<td>Caras (2019)</td>
<td>40 L2 Spanish (Novice)</td>
<td>The ser versus estar, the preterit versus imperfect</td>
<td>-Direct -Indirect – -Indirect + -Control</td>
<td>Think-aloud protocols</td>
<td>No difference among treatment groups in accuracy. Direct and indirect group processed at low level. Indirect + group processed at intermediate level.</td>
</tr>
<tr>
<td>Cerezo et al. (2019)</td>
<td>46 ESL (Intermediate)</td>
<td>All error types</td>
<td>-Direct -Indirect-with-codes -Control</td>
<td>Written languaging</td>
<td>Both treatment groups performed better than the control group, but there was no statistically significant difference between groups. Direct group processed at high level. Indirect + group processed at low level.</td>
</tr>
</tbody>
</table>

"Indirect" refers to indirect WCF without metalinguistic explanations. "Indirect +" refers to indirect WCF with metalinguistic explanations.
2.2.4 Summary

Whether and how different WCF types are effective has received a lot of attention in the literature. Researchers have examined focused WCF in specific contexts, such as only looking at English articles (e.g., Sheen, 2007) as discussed above. Researchers have also explored a wide range of errors under unfocused versus focused WCF, reporting more benefits for focused over unfocused WCF (e.g., Shintani et al., 2014). Besides error range, many studies have compared direct WCF, indirect WCF, and indirect WCF with metalinguistic explanations. Some researchers have argued that direct WCF is more effective (e.g., Karimi, 2016; Benson & Dekeyser, 2018, Cerezo et al., 2019), while many researchers have argued that indirect WCF is better (e.g., Lalande, 1982; Ferris, 1995; Bitchener & Knoch, 2008; Hartshorn et al., 2010). Kang and Han’s (2015) meta-analysis on WCF demonstrated that the results are inconclusive with regards to which type, when, what context, how and why feedback is most beneficial.

In an attempt to draw on the strengths of the WCF studies discussed above, this study synthesizes its research methods following Caras (2019) to reflect a realistic classroom environment’s unfocused WCF, while using two common and specific linguistic targets to measure changes in accuracy. In terms of feedback types, from a pedagogical perspective, it is very practical to provide direct WCF since it is the common practice of classroom teachers and many focused WCF studies to provide the empirical backing for the practice. Meanwhile, unfocused WCF studies have found that indirect WCF is more beneficial than direct, suggesting that there is room to challenge some of these preconceptions among practitioners. Note, however, that indirect WCF may comprise one of two treatments: one that only marks the error and one that is accompanied with metalinguistic explanation. Indirect WCF without metalinguistic explanation may make the feedback difficult for students to understand. For
example, one participant from Zheng and Yu’s (2018) study explained how they struggled to derive meaning from indirect feedback:

I was stuck at where the teacher underlined a sentence and put a question mark in the margin. I guessed it meant there was an error or some errors in the sentence, but I could not decide where exactly it was. So, I did not make any modifications to that sentence in my second draft.

This study does not use indirect WCF without metalinguistic explanations because there appears to be little support for it in the literature. Zheng and Yu (2018) recommended that teachers should intentionally explain written corrective feedback to low-proficiency students to reduce confusion and increase their cognitive engagement. It is more reasonable to adopt the indirect-WCF-with-metalinguistic-explanations method of including metalinguistic explanation to give students more guidance on what the error was and how they can correct it by drawing on the rich metalinguistic explanation and examples. This study compares both direct and indirect WCF with metalinguistic explanations to further contribute to the understanding of WCF effectiveness.

2.3 Processing Written Corrective Feedback

Current WCF research can better establish links between learning outcomes and WCF by addressing limitations in the study designs. Limitations include situating the studies within an unrealistic laboratory learning context or investigating only a small range of target languages and forms. More importantly, these studies have only rarely employed process measures to confirm that participants indeed paid attention to and processed the different types of WCF (Leow, in press). In most WCF studies, there is no evidence to demonstrate how participants processed the
different types of feedback provided on their errors. Investigation of WCF’s cognitive processing is clearly needed, analyzing how learners notice and react to the feedback, and how learners correct the error and modify output (Bitchener, 2018; Manchón, 2018; Caras, 2019; Cerezo, Manchón, & Nicolás-Conesa, 2019). This new line of research could shed light on the effectiveness of different types of WCF.

2.3.1 Theoretical Framework for Processing Feedback

Many empirical WCF studies have been conducted within the cognitive-interactionist domain and often cite the Noticing Hypothesis (Schmidt, 1993), the Interaction Hypothesis/Approach (Long, 1996; Mackey & Gass, 2006), the Output Hypothesis (Swain, 1985, 2005), and Skills Acquisition Theory (DeKeyser, 2015), and the Model of the L2 Learning Process in Instructed Second Language Acquisition (2015) as key theoretical underpinnings for WCF’s effectiveness and they serve as important concepts for understanding the cognitive processing of feedback.

The Noticing Hypothesis (Schmidt, 1995) draws from cognitive psychology’s premise that the L2 learner has a limited capacity to process information and selects certain information in the input for focal attention. Reflecting on his experience learning Portuguese (Schmidt & Frota, 1986), Schmidt (1995, p. 29) defined “noticing” as “conscious registration of the occurrence of some event,” whereas “understanding” is to imply recognition of a general principle, rule, or pattern. Noticing refers to surface level phenomena and item learning, while understanding refers to “deeper levels of abstraction related to (semantic, syntactic, or communicative) meaning, system learning.” In other words, if a learner wants to learn any linguistic feature of the target language, such as pronunciation, tones, grammar rules, or pragmatics, the feature in the target language input must be noticed by the learner. The learner
must pay focal attention to identify the feature, and this minimal focal attention is noticing, which involves a low level of awareness. If the learner goes beyond noticing and analyzes the underlying rules, this reaches awareness at the level of understanding. Many studies have found that while WCF triggered both noticing and understanding, benefits for writing accuracy were usually associated with the higher level of awareness (understanding) (e.g., Qi & Lapkin, 2001, Suzuki, 2017). Leow (2015) explained that the Noticing Hypothesis provides only a coarse-grained perspective of processing, given that merely noticing L2 data in the input is not enough for achieving language learning. Other variables associated with the process of noticing should be acknowledged when considering how new information is processed. For example, elaboration and evidence should be presented to account for whether the noticed target feature (or feedback) was processed further, and whether the learner internalized it into the internal system and produced the feature later. Additional frameworks, below, may help address some of these criticisms.

The Output Hypothesis (Swain, 1985, 2005) emphasizes the role of comprehensible output, drawing from Swain’s observations of student performances in L2 French immersion classrooms in Canada. She found that students had unbalanced performance in listening, reading, and speaking. Specifically, students reached near native comprehension ability in listening and reading, but their speaking ability did not match their listening and reading abilities. Therefore, Swain argued that providing students more opportunities to produce the L2 (such as speaking and writing) would increase their productive ability. In the Output Hypothesis, Swain made three major claims: 1) The opportunity to produce the target language output would push learners to identify potential learning gaps between their current interlanguage system and the target language; 2) The output opportunity would allow learners to try new rules and structures, and
then the learner would figure out whether their hypotheses for the new rules were correct or not; 3) the output opportunity would guide learners to analyze the mismatches identified in their production of the target language. At the same time, Swain also claimed the importance of “languaging,” a type of meta-talk, where the learner uses language to talk about the language learning experience. Many researchers adopted the Output Hypothesis to explain the writing process and collaborative metalinguistic discussion between L2 learners and peers or native speakers. For example, Wigglesworth and Storch (2012) focused on metalinguistic written feedback discussion with native speakers and argued that the feedback offered an opportunity for learners to identify the mismatches between their production and the target forms. The feedback would lead learners to analyze the underlying rules and experiment on their hypothesis of the rules, leading to L2 development.

Skills Acquisition Theory (DeKeyser, 2015) defines three stages for learners to develop a new skill: declarative knowledge, proceduralization, and automatization. Declarative knowledge refers to “knowledge about a skill without trying to use it, [that] may be acquired through instruction (p. 97).” Learners’ action on declarative knowledge is proceduralization, where they establish the ability to apply the knowledge in some form. Automatization refers to when “relevant behavior was consistently displayed with complete fluency and spontaneity, rarely showing any errors (p. 98).” He claimed that the explicit L2 knowledge needed to be proceduralized over time through meaningful, contextualized practice, ultimately being converted into implicit knowledge. Researchers have asserted that WCF offers learners opportunities to practice the target language more accurately until learners convert declarative knowledge into procedural knowledge. For example, Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause, and Anderson (2010) experimented with “dynamic feedback” practice where
participants received WCF and completed revisions until the drafts were error free. Given the many meaningful practice opportunities, participants improved their writing accuracy to the stage of proceduralization.

Leow’s (2019) Feedback Processing Framework (Figure 2), which provides the theoretical underpinning of WCF processing in this study, is based on Leow’s (2015) Model of the L2 Learning Process in Instructed Second Language Acquisition. The framework provides a fine-grained cognitive explanation of how L2 writers process L2 feedback along several stages of the L2 learning process.

![Feedback processing framework based on Leow’s (2015) Model of the L2 Learning Process in Instructed Second Language Acquisition](image)

Figure 2. Feedback processing framework based on Leow’s (2015) Model of the L2 Learning Process in Instructed Second Language Acquisition

According to Leow (forthcoming), L2 writers process WCF in the following stages:

(1) learners receive L2 information via WCF based upon their own L2 written product;

(2) learners cognitively process the feedback in relation to their prior knowledge;

(3) the further processed information either reinforces or restructures their interlanguage lodged in the knowledge system; and
(4) when there is a new opportunity to produce the output, learners may produce old output or new or modified output dependent upon how the feedback was or was not cognitively processed. This framework includes consideration of learners’ depth of processing, “the relative amount of cognitive effort, level of analysis, elaboration of intake together with the usage of prior knowledge, hypothesis testing, and rule formation employed in decoding and encoding some grammatical or lexical item in the input,” (Leow, 2015, p. 204). In other words, if a learner processes certain feedback information and their cognitive effort is greater during processing, for example, using the linguistic resources to understand the information, then the learner is more likely to retain the information that they processed (please see examples in Table 4.)
Table 4. *Examples of Three Levels of Depth of Processing (DoP)*

<table>
<thead>
<tr>
<th>Low DoP</th>
<th>Medium DoP</th>
<th>High DoP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>Shows no potential for processing target</td>
<td>Comments on target in relation to meaning</td>
</tr>
<tr>
<td><strong>Descriptors</strong></td>
<td>Reads feedback quickly</td>
<td>Spends more time processing target</td>
</tr>
<tr>
<td></td>
<td>Says they aren’t sure what is wrong</td>
<td>Makes comments that indicate some processing of target, refers to a previous instance with the same target</td>
</tr>
<tr>
<td></td>
<td>Avoids the correction</td>
<td>Some level of cognitive effort to process target</td>
</tr>
</tbody>
</table>

2.3.2 Procedures for Assessing Processing

To understand how learners process L2 data, researchers (e.g., Calderón, 2013; Godfroid, Boers, & Housen, 2013; Leow et al., 2008; Leung & Williams, 2011; Morgan-Short et al., 2012; Qi & Lapkin, 2001; Rott, 2005) have employed different procedures (e.g., eye-tracking, think aloud protocols, reaction times) to gather data on cognitive processes (e.g., attention, noticing, awareness) employed by L2 learners. This processing-oriented line of research has also been adopted recently to shed light on the effectiveness of different types of WCF and achieve a better understanding of why certain results are observed, whether in the laboratory or classroom.
Understanding the different methods is important for interpreting their application in WCF research summarized in the subsequent sections of this chapter. There are currently four major data elicitation procedures employed to address how L2 writers process language information, namely, language related episodes (LRE) (Coyle, Cánovas-Guirao, & Roca de Larios, 2018); written languaging (e.g., Cerezo et al., 2019; Suzuki, 2012, 2017), concurrent think-aloud protocols (e.g., Caras, 2019; Park & Kim, 2019; Storch & Wigglesworth, 2010), stimulated recall interviews (e.g., Qi & Lapkin, 2001), and eye-tracking (Gilabert et al., 2016). These procedures are summarized in Table 5 and discussed in greater detail below.
<table>
<thead>
<tr>
<th>Method</th>
<th>Definition/Description</th>
<th>Strengths</th>
<th>Weaknesses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language related episodes</td>
<td>(Swain &amp; Lapkin, 1995) “any segment of the protocol in which learners either spoke about a language problem they encountered while writing and solved it either correctly or incorrectly, or simple solved it without having explicitly identified it as a problem.”</td>
<td>Introspective</td>
<td>Passive, observational</td>
</tr>
<tr>
<td>Written languaging</td>
<td>Swain (2006) “languaging” as “the process of making meaning and shaping knowledge and experience through language</td>
<td>Introspective</td>
<td>Reactivity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Time consuming</td>
</tr>
<tr>
<td>Think-aloud protocols</td>
<td>Leow (2015)</td>
<td>Introspective</td>
<td>Reactivity</td>
</tr>
<tr>
<td></td>
<td>Verbalizing the thought processes while a learner is performing a task</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stimulated recall interviews</td>
<td>Mackey &amp; Gass (2016)</td>
<td>Retrospective</td>
<td>Memory decay</td>
</tr>
<tr>
<td></td>
<td>learners are showed a video of doing a task and asked to talk about what they were thinking at the time that they were doing the task.</td>
<td></td>
<td>Misattribution</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Double exposure</td>
</tr>
<tr>
<td>Eye tracking</td>
<td>Javal (1879)</td>
<td>Not intrusive</td>
<td>Attention-focused</td>
</tr>
<tr>
<td></td>
<td>eye movement behavior using visual observation of the eyeballs.</td>
<td></td>
<td>Mental process must be inferred.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Equipment required</td>
</tr>
</tbody>
</table>

Language related episodes are a method for segmenting and analyzing dialogues, generally between learners. They were first defined by Swain and Lapkin (1995, p. 378), as “any segment of the protocol in which learners either spoke about a language problem they encountered while writing and solved it either correctly or incorrectly, or simply solved it without having explicitly identified it as a problem.” In other words, a language related episode is any part where learners reflected on their language use in their dialogue. This method, when used by itself, can be uneven because it relies on passively observing the episode and recording it, while inferring the meaning separately. In some cases, language related episodes are used as a
method of extracting segments of a dialogue for further assessment through other measures, such as think a louds, that can explore what the learner considered during the segment. Some studies have adapted language related episodes to the WCF context (e.g., Qi & Lapkin, 2001; Coyle et al., 2018).

Written languaging is a method of recording, in writing, the internal monologue or thought process while absorbing new information. Chi, Bassok, Lewis, Reimann, and Glaser (1989) introduced the concept of written languaging as “the activity of explaining to oneself in an attempt to make sense of new information” (p. 164) and as “a knowledge-building activity that is generated by and directed to oneself” (p. 165). In second language acquisition, Swain (2005) further defined “languaging” as “the process of making meaning and shaping knowledge and experience through language" (p. 98) and she argued languaging is a powerful tool to understand the role of language in mediating cognition. While producing high utility data at minimal cost to the researcher, this method bears a high risk of reactivity given what is likely a high cognitive load in writing one’s thoughts separately from conducting the task at hand, and the potential to cause greater depth of processing through increased introspection. It can also be very time consuming for the participant as they must move from correction to written languaging during the revision stage while writing.

Think aloud protocols are another type of introspective data elicitation procedure and are used to collect data on the thought processes employed by learners as they perform a task (Mackey & Gass, 2016). Concurrent think aloud protocols originated in the fields of psychology and classical philosophy where participants dictated their thoughts for researchers (Ericsson & Simon, 1993). Think alouds can provide insights into the processes, processing and strategies employed by learners as they interact with L2 data at a lower cognitive cost than written
language, though at a higher cost to the researcher who must generally transcribe the recordings (Leow, Gray, Marijuan, & Moorman, 2014). However, think aloud protocols are an intrusive procedure that may cause reactivity in the participants by potentially distracting them from the task at hand or encouraging greater processing through the think-alouds’ introspective monologue. Reactivity means “the act of simultaneous reporting might serve as an additional task altering the very thinking processes it is supposed to represent and keep intact (Wilson, 1994). In other words, it is possible that thinking aloud might change the primary purpose of the task or add additional cognitive load on learners. Bowles (2010) conducted a meta-analysis on studies published up to 2008 and concluded that think aloud protocols did not have a consistent reactivity effect.

Yanguas and Lado (2012) investigated 37 heritage learners of Spanish and asked them to perform a semi-guided writing task in two groups (think-aloud protocols and no-think-aloud protocols). They explored the validity of employing think aloud protocols in heritage language writing, and found that the think-aloud protocols group outperformed the control group in both accuracy and fluency. They asserted that think aloud protocols should be implemented with caution. In their view, implementing think-aloud protocols during the written tasks may trigger reactivity and facilitate the learning process. Leow et al. (2014) also underscored that while researchers must still be cognizant of the level of intrusiveness, the presence of reactivity may vary depending on the experimental task types.

Stimulated recall interviews are a type of retrospective data collection procedure and occur separately after the completion of the task (Mackey & Gass, 2016). This eliminates risk of reactivity, but introduces other factors. A stimulated recall, also called a retrospective verbal protocol, asks learners to reflect on their thought processes, usually while being shown a video of
themselves doing tasks (or alternatively a written transcript). Researchers ask learners to talk about what they were thinking at the time that they were doing the L2 task, and may guide learners to focus on specific points. One major critique of stimulated recall is that the learner might misattribute their thoughts, and verbalize what they are thinking at the time of the recall interview and not what they were thinking during the task (Mackey & Gass, 2016). Another limitation may be veridicality, that is, participants may forget what they were thinking at that time (memory decay) and have to fill in the gaps during the interview. Finally, there is the limitation of exposing participants to the L2 data for a second time, which may impact their responses during the after-exposure stage.

Lastly, eye tracking is a method first proposed by Javal (1878), who examined eye movement behavior using visual observation of the eyeballs by the researcher (Mackey & Gass, 2016). Modern usage often relies on computerized, camera-based eye tracking, making it a passive tool for data collection (to the extent that specialized equipment does not distract the participant). Researchers have argued that overt attention (as shown by the exact position of the eyes, as measured by the movement of light reflecting of the cornea and pupil) and covert attention (mental focus) are closely related (e.g., Smith 2012; Godfroid & Uggen, 2013; Godfroid, 2019). The eye-tracking procedure offers an online record of the processing in learner’s ‘natural’ reading behavior without contaminating the task. However, researchers have suggested that eye-tracking procedures still do not reveal learner’s mental processes such as depth of processing, linguistic form understanding, and interlanguage restructuring (Winke, 2013).
2.3.3 Process-oriented Studies of Written Corrective Feedback

The major studies that have applied these process measures to explore WCF are summarized in Table 6, below. Some of the studies apply multiple methods. The remainder of this subsection examines these studies and their results in greater detail.

Qi and Lapkin (2001) examined the role that noticing plays on composing and on the processing of the “reformulation” of an L2 composition task—researchers corrected all of the syntactic and morphological errors, and removed any problems in stylistics and logical sequencing. Two adult Chinese English as a second language participants (intermediate level) wrote a first draft, and compared their draft with a reformulation while completing a think aloud protocol, then revised the first draft. The researchers and participants completed stimulated recall interviews after their revisions and authors found that language-related noticing had a close relationship with development of L2 writing. More importantly, they found two levels of noticing: perfunctory noticing (noticing only) and elaborate noticing (noticing and providing reasons). They found that the level of noticing occurred at the second stage with reformulation and at the final revision stage. Qi and Lapkin suggested that the elaborate noticing in language related episodes that occurred during the written corrective stage was important to improve the final draft. They argued that teachers should work with low-proficiency learners to improve the noticing quality; however, it is difficult to generalize their findings to other contexts since only two participants were examined in this study. A higher n-size in future research would make these findings more compelling.

Sachs and Polio (2007) examined 15 intermediate ESL learners’ writing accuracy after receiving two types WCF: direct versus reformulations. Think-aloud protocols were employed to analyze participants’ awareness. Sachs and Polio found the direct WCF group performed
significantly better than the control and reformulation treatment groups in their revisions. Think-aloud data suggested noticing of WCF promoted writing accuracy in the revisions.

Wigglesworth and Storch (2012) explored the language learning benefits from processing feedback in pairs and used think-aloud protocols to understand two pairs of peer English as a second language learners’ processing of WCF. They found indirect WCF had a higher engagement than direct feedback, and led learners to collective scaffolding—both learners used their linguistic resources and reached resolutions. However, the deeper indirect feedback processing did not lead to a lasting effect on the delayed posttest. They found direct feedback produced better accuracy on the delayed posttest due to different feedback strategies. Learners who received direct WCF used a memorization method and certain errors such as spelling and punctuation did not need deeper processing.

Coyle, Cánovas-Guirao, and Roca de Larios (2018) also investigated the language learning benefits of processing feedback in pairs, but not with peers. Eight ESL primary schoolers were paired with their teachers to discuss WCF and improve writing accuracy. Coyle et al. found these 10-12 year old participants outperformed the control group on writing accuracy. They also provided rich data to explain how these 8 young ESL learners followed different sequential trajectories in their feedback processing and writing. They found that the ideal routine was to notice a problem, see the solution in the direct WCF, process the solution semantically and syntactically, use the new information to create a hypothesis, and finally add the new information to their internal language system. Their data indicated that many participants did not follow this ideal trajectory. Many of them did not notice the solution or only partially noticed the solution in the direct written feedback. The use of the ideal trajectory was related to L2 writing accuracy and knowledge development. Overall, Coyle et al.’s (2018)
procedure may be too intrusive to allow for generalizable results. Not only were participants provided with WCF, but participants had discussions with their teachers. It is quite possible that the noticing of the feedback and writing accuracy improvements found in the study were caused by teachers’ explanation in the discussion sessions.

Suzuki (2012) investigated the processing of written feedback from a sociocultural-psychology perspective using written languaging. 24 intermediate L2 English participants wrote an essay in 30 minutes, wrote explanations in L1 Japanese about the direct WCF that they received, and revised their essays. Suzuki found that learners’ languaging about their direct WCF in the first essay helped them successfully understand and correct errors during immediate revision. Suzuki’s 2012 study did not focus on depth of processing or levels of awareness of feedback. However, Suzuki (2017) conducted a follow-up study on the relationship between the levels of awareness and the accuracy of immediate revisions. Processing direct WCF was analyzed in terms of three “levels of awareness” that reflect noticing the correction, understanding the reasons for it, and uncertainty about the intent of the feedback. The study reported that learners mostly engaged in the two levels of noticing, which are noticing only and noticing with understanding. Both levels of awareness had similarly high accuracy rates in their revisions.

Park and Kim (2019) explored indirect WCF without metalinguistic explanations, different proficiency levels, and participants’ depth of processing as they self-corrected their errors in the writing tasks. 24 participants at two proficiency levels (elementary and intermediate) wrote two different writing tasks and completed think-aloud protocols while looking at their indirect WCF. They reported that participants from both of the two proficiency groups could self-correct with indirect WCF. The depth of processing data showed that higher
levels of processing resulted in better self-correction. However, the participants from the low-
proficiency group experienced difficulty in accurately identifying the error type. Park and Kim
explained that if a participant only had partial or inaccurate prior knowledge of the target error, it
would be difficult for the participant to correct the error even though participants reached a high
level of depth of processing. Therefore, they argued the teacher should not only give indirect
WCF by underlining the error but also provide additional instruction or examples.

Other researchers have argued that direct WCF requires little depth of processing
compared to indirect WCF (Shintani & Ellis, 2013); however, there have been only three studies
that investigated indirect WCF with metalinguistic explanations versus direct WCF. Firstly,
Storch and Wigglesworth (2010) compared the level of processing of indirect with metalinguistic
codes versus direct WCF in participants’ think aloud recording and the reformulation stage.
There were two levels of processing in their data: “extensive engagement,” including comments
demonstrating metalinguistic awareness of the WCF, and “no engagement,” involving simple
acknowledgement of the feedback and no further processing. They found that more language
related episodes and a higher level of engagement occurred with the indirect feedback with
metalinguistic codes.

In a similar study, Caras (2019) employed think-aloud protocols in a comparison of
different types of WCF. Fifty-one participants in L2 Beginning Spanish were randomly assigned
to one of three unfocused written corrective feedback treatment groups (direct, indirect WCF
without metalinguistic explanations, and indirect WCF with metalinguistic explanations), and a
control group. The indirect WCF with metalinguistic explanations group was provided richer
feedback, including additional explanations with examples along with the error code.
Participants finished a composition and revised it after the WCF treatment, before revising the
original composition one week after writing the original composition. Caras found that the majority of the direct and indirect WCF without metalinguistic explanations groups processed target linguistic items at a low level of processing. However, the indirect WCF with metalinguistic explanations group processed the targets at medium and high levels. The WCF types had no differential effect on accuracy scores over time. The effects of think-aloud protocols were then examined by measuring the success of immediate subsequent text revisions. There were three major findings: First, unfocused WCF was not automatically processed and understood, even when participants were thinking aloud about their own errors. Second, if participants understood both the error and the WCF, thinking aloud on WCF promoted participants’ successful correction in the immediate revision stage. Last, participants’ depth of processing of feedback was related to their decision making about how to understand the WCF and what to do with the feedback.

Finally, Cerezo et al. (2019) employed written languaging to investigate the relationship between depth of processing on WCF and written accuracy. Forty-six intermediate undergraduates were divided between direct WCF and indirect with metalinguistic codes treatments, and a control group. The participants completed a three-stage writing task comprised of writing, receiving and processing WCF, and revision. Both feedback groups performed better than the control group, but with statistically significant differences in the amount of successful corrections and errors left uncorrected in the revised texts. Surprisingly, the study found that those receiving direct WCF demonstrated greater depth of processing, indicative of awareness at the level of understanding and explaining errors. In comparison, the indirect-with-metalinguistic-codes group tended to only report a noticing level. By finding that WCF did not automatically
get processed fully and accurately, the study highlighted the importance of further research into how WCF is perceived and processed by learners.

2.3.4 Summary

In most WCF studies, scant attention has been paid to WCF’s cognitive processing. Only a few studies have demonstrated how participants processed the different types of feedback provided on their errors. This new line of research has employed four major data elicitation procedures to address how L2 writers process language information, namely, language related episodes (LRE) (Coyle, Cánovas-Guirao, & Roca de Larios, 2018); written languaging (e.g., Cerezo et al., 2019; Suzuki, 2012, 2017), concurrent think-aloud protocols (e.g., Caras, 2019; Park & Kim, 2019; Wigglesworth & Storch, 2010), and stimulated recall interviews (e.g., Qi & Lapkin, 2001).

Table 6 provides a review of current studies examining depth of processing of WCF and it reveals inconclusive findings. First, there is little data or literature that examines how participants processed different types of WCF. For example, the majority of studies (Qi & Lapkin, 2001; Suzuki, 2012, 2017; Coyle et al., 2018) examined direct feedback, and only Park and Kim (2019) examined indirect feedback. Wigglesworth and Storch (2012) compared direct and indirect WCF, however, their participants completed writing tasks in pairs and each pair had a 15-minute discussion on their WCF. Therefore, it is difficult to directly link the observed learning outcomes to WCF due to methodological limitations.

Second, the three studies that compared direct WCF and indirect WCF with metalinguistic explanations had inconsistent results. Wigglesworth and Storch (2012) found indirect WCF with metalinguistic explanations was processed at a high level of processing. Caras (2019) provided supporting evidence for that conclusion, finding that the indirect-with-
metalinguistic-codes (no explanations) group also processed the feedback at high and intermediate levels. However, Cerezo et al.’s (2019) written languaging data provided a contrasting finding, with the indirect-with-metalinguistic-explanations group processing at a low level. The inconsistent results could be explained by the number of participants, level of proficiency, or how the two types of WCF were operationalized. Wigglesworth and Storch (2012) only had four ESL participants and these participants completed the writing task and feedback discussion in pairs. Even though both Caras (2019) and Cerezo et al. (2019) had more than 40 participants, the two studies had different proficiency levels and feedback procedures. Caras’ participants were beginners and they were provided both indirect WCF and indirect WCF with metalinguistic explanations. Only the indirect WCF with metalinguistic explanations group processed at a high level. In Cerezo et al. (2019), the participants were intermediate L2 writers and the indirect WCF was only metalinguistic codes without explanations. It was possible that their participants did understand the metalinguistic codes and processed the feedback at low level.

Given the inconsistent findings on the depth of processing for direct WCF versus indirect WCF with metalinguistic explanations found in Wigglesworth and Storch (2012), Cerezo et al. (2019) and Caras (2019), more evidence of WCF cognitive processing is needed, especially in order to analyze how learners notice and react to different types of WCF, and how learners correct their errors and modify their output (Bitchener, 2018; Manchón, 2018; Caras, 2019; Cerezo et al., 2019). In an attempt to draw on the strengths of the WCF studies discussed above, the current study synthesized research methods from Caras (2019) and employed think-aloud protocols to explore how participants processed different types of WCF in an attempt to explain why certain types of feedback were more effective.
Table 6. Overview of Studies Employing Language Related Episodes, Written Languaging, and Think-Aloud Protocols

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Processing Task</th>
<th>WCF Types</th>
<th>Process measures</th>
<th>Depth of Processing (DoP) Or Level of Noticing Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qi &amp; Lapkin (2001)</td>
<td>2 ESL (Intermediate)</td>
<td>WCF processing</td>
<td>-Reformulation</td>
<td>Think aloud</td>
<td>Perfunctory noticing (noticing only) Elaborate noticing (noticing and providing reasons) helped accuracy improvement.</td>
</tr>
<tr>
<td>Sachs &amp; Polio (2007)</td>
<td>15 ESL (Intermediate)</td>
<td>WCF processing</td>
<td>-Reformulation</td>
<td>Think aloud</td>
<td>Direct group outperformed reformation and control group on writing accuracy. There were two level of noticing: Noticing without reasons Noticing with provision of reasons &amp; use of metalanguage</td>
</tr>
<tr>
<td>Wigglesworth &amp; Storch (2012)</td>
<td>4 ESL (Intermediate)</td>
<td>WCF processing (collaborative)</td>
<td>-Direct -Indirect+</td>
<td>Think aloud</td>
<td>Indirect had a higher engagement than direct. Direct had a higher accuracy in delayed posttest.</td>
</tr>
<tr>
<td>Suzuki (2012, 2017)</td>
<td>24 ESL (Intermediate)</td>
<td>WCF processing</td>
<td>-Direct</td>
<td>Witten Languaging</td>
<td>There were two levels of noticing in the WCF processing: Noticing the correction only and noticing with reasons. Both helped to improve writing accuracy in the revision.</td>
</tr>
<tr>
<td>Caras (2019)</td>
<td>41 L2 Spanish (Novice)</td>
<td>WCF processing</td>
<td>-Direct -Indirect -Indirect + -Control -Direct</td>
<td>Think aloud DoP</td>
<td>No significant differences among groups. Indirect + had higher level of DoP. Direct and indirect had low level of DoP.</td>
</tr>
<tr>
<td>Coyle et al. (2018)</td>
<td>8 ESL (Novice)</td>
<td>WCF processing (collaborative)</td>
<td>-Direct</td>
<td>Think aloud</td>
<td>Primary school students had different feedback processing trajectories. Deeper processing led to higher accuracy.</td>
</tr>
</tbody>
</table>
Table 6. (Cont.)

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Processing Task</th>
<th>WCF Types</th>
<th>Process measures</th>
<th>Depth of Processing (DoP) Or Level of Noticing Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerezo et al. (2019)</td>
<td>46 ESL (Intermediate)</td>
<td>WCF processing</td>
<td>-Direct</td>
<td>Written languaging</td>
<td>Direct group had high level of depth of processing. Indirect processed at a low level.</td>
</tr>
<tr>
<td>Park &amp; Kim (2019)</td>
<td>24 ESL (Novice and intermediate)</td>
<td>WCF Processing</td>
<td>-Indirect</td>
<td>Think aloud</td>
<td>Both proficiency level learners accurately corrected more than one-third of their errors. The think-aloud results showed that the majority of underlined errors were processed either at the highest or the lowest level of processing</td>
</tr>
</tbody>
</table>

*“Indirect” means indirect WCF without metalinguistic explanations. “Indirect +” means indirect WCF with metalinguistic explanations. Reformation means native speaker rewrite the sentence without changing the meaning. Cerezo et al. (2019) provides indirect WCF with only metalinguistic codes.*
2.4 Feedback Engagement and Teacher Perspectives on Written Corrective Feedback

There is a growing body of research on student engagement with written corrective feedback that explores WCF complexity and individual differences (e.g., Hyland, 2003; Ferris, Liu, Sinha, & Senna, 2013).

Ellis (2010) proposed a framework for investigating student engagement (both oral and written) from three perspectives: affective, behavioral, and cognitive (see Figure 3). Affective engagement refers to students’ emotional reaction to receiving WCF, including how they feel about themselves how they feel towards the feedback, and how they feel toward the teacher as the provider of the feedback (Zheng & Yu, 2018). The behavioral perspective refers to how students handle the WCF, for example, by incorporating WCF in the revision (Hyland, 2003) or using strategies in the revision (Ferris et al., 2013). The cognitive perspective refers to students’ cognitive effort expended to process WCF. Cognitive effort or investment can manifest in greater depth of processing (Leow, 2015; Caras, 2019). Some studies also examined cognitive efforts...
along the lines of Schmidt’s (1990) awareness at the level of noticing and at the level of understanding (Suzuki, 2012, Cerezo et al., 2019).

All three perspectives are important to fully understand student engagement, yet they have received little attention in second language acquisition research (Zheng & Yu, 2018; Han & Hyland, 2019). To the author’s knowledge, only five studies have examined participants’ engagement with WCF. Han and Hyland (2015) investigated the engagement of four intermediate L2 English students receiving WCF, applying a qualitative lens. They discovered that the four students’ engagement varied because of different beliefs, different learning experiences, and the different context of receiving and processing the feedback. They analyzed the participants’ drafts and transcriptions of interviews, verbal reports, and teacher-participant discussion. They found all four participants made a wide range of errors (e.g., in word choice and sentence structures), but the teacher usually gave them WCF selectively and implicitly, such as indirect by underlining and circling. Sometimes, the teacher offered revision clues or explanations in the margin. Out of four case study participants, “Ying” (pseudonym), who made the most progress, was very motivated and highly engaged with teacher’s feedback. Ying produced metalinguistic explanations for 10 out of 15 errors and processed these errors at the level of understanding in her verbal report. Another participant, “Song” (pseudonym), was the least motivated participant in this study. In the interview after the feedback treatment, she indicated that she just wanted to pass the course. In her verbal report, she stated she was emotional, upset, and very disappointed. When she saw her feedback, she was unable to concentrate and felt uncertain about revision. She noticed 19 of 21 errors, but only processed them at a minimal level, and she did not correct the errors based upon the feedback. An additional participant, “Dai”, was moderately motivated, and her understanding of teacher feedback was also at a low level because
of limited metalinguistic knowledge of errors. Lastly, “Lin” barely made any progress after receiving feedback because he perceived it as a wake-up call to work on his English test and a warning of his weakness in grammar. In his verbal reports, Lin didn’t pay attention to feedback and error corrections, and he was satisfied with his first draft. These four cases revealed the complex relationship between participants’ beliefs and their feedback engagement. The study highlighted the context and individual differences of participants’ feedback engagement. These four cases also added supportive evidence for Caras (2019) that participants who processed the WCF at the level of understanding would have a successful revision.

Zhang (2017) studied an L2 English university student’s engagement with computer-generated WCF. He found that the student could actively engage in feedback processing but the emotional engagement was influenced by the writing score and the overall behavioral pattern was not clear. Mahfoodh (2017) focused on how students responded emotionally to WCF. Mahfoodh argued that both positive (surprise, happiness) and negative (dissatisfaction, frustration) emotional responses can affect learners’ understanding and uptake of WCF. Zheng and Yu (2018) examined how twelve low-proficiency L2 English learners engaged with WCF. They found that students responded positively in the emotional aspect of engagement, but students’ low English proficiency negatively influenced cognitive and behavioral engagement with WCF. Students did not improve in their revisions, and the study suggested more metalinguistic explanations should be added for beginner students to increase engagement.

Lastly, Han and Hyland (2019) investigated two advanced L2 English learners’ discrete emotional reactions toward WCF through semi-structured individual interviews and retrospective verbal reports. Han and Hyland selected two extreme cases (participants: Du and Hong) based upon their performance level in their classes. Du made 36 errors in a 799-word essay and Hong
made 39 errors in a 411-word essay. Du was a highly motivated student and had strong metalinguistic awareness. In her verbal report, she was positive and described being in a problem-solving mode when she received WCF. As for the other participant, Hong was not interested in English learning and struggled with English for years. When Hong saw his teacher’s feedback, he tried to use humor to disguise his embarrassment and anxiety. Hong noticed his errors, but he downplayed them as “silly errors” and explained that he was careless, but not incompetent. It was interesting that Han and Hyland found both high-achieving students and low-motivation students had rich and dynamic emotional reactions to WCF. These two cases supported Mahfoodh (2017)’s finding that WCF can trigger both positive, negative, and neutral emotions. Most interestingly, the study found that negative emotions elicited from WCF are not necessarily destructive—the negative emotion might be temporary and ultimately overridden by positive and neutral feelings as they push through. Further, negative emotions increased students’ motivation and performance in the revision stage. Surprisingly, positive emotions might reduce mental effort in the future drafts.

Examining emotions from a holistic perspective, Dewaele and MacIntyre (2019) conducted a mixed-methods study that was based on a questionnaire of 750 foreign language learners around the world. They assessed quantitative results from the Likert scale responses, and qualitative data from descriptions of classroom episodes in which participants experienced foreign language emotions and classroom anxieties. Dewaele and MacIntyre found foreign language emotions and foreign language anxieties were separate, independent dimensions. They found that language emotions were mostly related to teacher-centered variables, and class anxieties were mostly predicted by the personality traits of the student. In other words, the findings showed teachers could influence learners’ emotions, while anxieties were mostly related
to learners themselves. Therefore, their study implies that teachers should focus on boosting positive emotions and not worry about managing student anxieties.

Teachers provide a variety of corrective feedback to their students in authentic classrooms, often with limited understanding of which type of feedback may be best suited to a given situation, task, or an individual. WCF not only plays a critical role in the writing course but recently has drawn much attention for its potential instructional role in the foreign language classrooms. Research has shown that WCF can be utilized as an instructional tool to improve students’ writing content, accuracy and complexity (e.g., Bitchener & Ferris, 2012; Yu & Lee, 2016; Riazi et al., 2018); however, there are challenges in the real-world implementation of WCF. First, studies on teachers’ perspectives on WCF show that teachers often view their role as “error hunters” and push themselves to focus on providing very comprehensive feedback that highlights every error in a document. They view WCF as sometimes unpleasant (Truscott, 2001) and very time-consuming (Ferris, 2003) because of this self-imposed dynamic. Given the large workload and time pressure common in the teaching profession, WCF might not always be accurate (Lee, 2004). Second, although research has indicated that providing WCF in a focused manner is more beneficial than trying to identify all types of errors (e.g., Bitchener, 2008; Bitchener & Knoch, 2008, 2010a, 2010b; Ellis, Sheen, Murakami, &Takashima, 2008; Sheen, 2007), these findings have not influenced most teachers’ classroom practice (Lee, 2018). One contributing factor may lie in the difficulty of translating findings from the laboratory environment where most WCF studies have taken place. Laboratory study designs bear few similarities to real classrooms, with researchers choosing to give focused WCF on one or two error types. This format is very different from the actual classrooms where teachers provide WCF on a range of error types. This is a key weakness regarding pedagogical implications.
derived from these studies (Storch, 2010), reflecting limited consideration for the perspectives of practitioners who may seek to integrate new methods in their curriculums.

Lee (2018) explored teachers’ perspectives on WCF and sought to inform and influence pedagogical methods in real classrooms. Through a survey of 90 EFL teachers in Hong Kong, Lee collected perspectives on WCF on some research-based recommendations. These EFL teachers argued that providing limited WCF would upset students, parents and school leaders who are accustomed to seeing comprehensive corrections. A change to focus on only one correction would be viewed as laziness or carelessness on the part of the teachers. In addition, targeting only one or a limited range of errors was found to not suit students’ learning needs since students want to improve on overall accuracy. Interestingly, Lee’s study found that 73% of teachers cared most about students’ positive engagement with WCF. These teachers hoped that their students would act upon the feedback, “grasp knowledge” and “develop a higher self-esteem” (Lee, 2018). Lee (2018)’s study points to an important issue in that only through understanding why and how students react and respond to WCF can teachers enhance the understanding of feedback effectiveness and apply research findings in an innovative way. Students should not be treated as passive receivers of feedback, but should be viewed as the active engagement center of WCF.

In summary, teachers may view WCF as an important instructional tool to improve students’ writing and learning, however, it is quite challenging to apply findings from WCF research due to the disparity between the research laboratory context and real classrooms. Furthermore, teachers’ pedagogical practices on WCF should be informed by a comprehensive understanding of students’ engagement with the material, as it is likely to effect outcomes. Using the affective, behavioral, and cognitive perspectives, studies have found students’ engagement
with WCF might be influenced by proficiency level, beliefs, learning experiences, emotional response, and feedback contexts. Students have rich and dynamic emotional reactions to WCF, and some negative emotions may actually help in the learning process. To further improve knowledge and practices on WCF, authentic classroom-based research, with greater ecological validity and a thorough understanding of students’ engagement are necessary.

2.5 Current Study

The current study examined the development of accuracy in L2 Mandarin Chinese writing in a task-based curriculum context. The literature on written corrective feedback has demonstrated clear benefits for SLA, yet has no clear answers on what types of feedback might be most useful in different contexts. Studies argued that focused WCF yields better results than unfocused (e.g., Ellis et al., 2008; Sheen et al., 2009), however, this difference might arise from a greater ability to target and measure how participants improved given that this research focuses precisely on the accuracy change of one or two target items. From teachers’ pedagogical perspectives, it is important for research to reflect realistic classroom environments by providing unfocused feedback. Teachers felt it was impractical to only focus on one or two linguistic targets, and that error range and selection was too complex to make informed decisions (e.g., Storch, 2010; Hyland & Wong, 2013; Lee, 2019).

In the line of unfocused WCF research, some researchers believed that direct WCF was more effective (e.g., Karimi, 2016; Benson & Dekeyser, 2018, Cerezo et al., 2019), while other researchers argued that indirect WCF was better (e.g., Lalande, 1982; Ferris, 1995; Bitchener & Knoch, 2008; Hartshorn et al., 2010). Kang and Han’s (2015) meta-analysis on WCF demonstrated that the results were inconclusive with regards to which type, when, what context, how and why feedback is most beneficial. It was insufficient at this point to continue to research
only the outcomes of different WCF treatments. Consideration must be given to the learner’s processing and the context in which feedback is provided (e.g., Bitchener, 2018; Ferris, 2006; Leow, 2019; Manchón, 2018). The next step in WCF research is to continue to gather more nuanced insights into how learners are processing the different types of WCF in order to offer explanations as to why conflicting results may be observed, as exemplified in studies that have employed concurrent data collection procedures (e.g., Suzuki, 2017; Bitchener, 2018; Manchon, 2018; Caras, 2019; Cerezo et al., 2019). Building on the use of process measures employed in previous studies, this study will examine accuracy alongside how L2 writers notice, understand, and act on WCF by addressing the following research questions:

2.5.1 Research Questions

1. What is the effect of different types of WCF (direct vs. indirect with metalinguistic explanations) on low-proficiency L2 Chinese writers’ accuracy with le and classifiers?

2. How do low-proficiency L2 Chinese writers process the two different types of WCF (direct vs. indirect with metalinguistic explanations) on their written production?

3. How do low-proficiency L2 Chinese writers engage with the two different types of WCF (direct vs. indirect with metalinguistic explanations)?

4. What are teachers’ perspectives about the two different types of WCF (direct vs. indirect with metalinguistic explanations)?
CHAPTER III: METHODOLOGY

This chapter describes the participants and the context of the program that hosted the study. The target linguistic items are then presented followed by procedures and materials. The last section is dedicated to detailing the study’s data collection techniques and data encoding methods.

3.1 Participants

Participants were drawn from 45 adult learners of Mandarin Chinese enrolled in an introductory online course that requires a weekly, 45-minute video-chat session with a teacher. This introductory Mandarin Chinese course is module-based and runs over 14 weeks. It is designed specifically for professionals who will live in China. All of the learners in this course self-reported their Mandarin Chinese ability within the range of ab initio beginners to the equivalent of ACTFL Novice (ACTFL, 2012). Participants agreed to join the current study as part of their course and were not offered additional compensation for their participation. All the participants were randomly assigned to one of two WCF treatment groups—1) direct, and 2) indirect with metalinguistic explanations. There is no control group in this study because the study was conducted in an authentic foreign language classroom and it would be unethical and impractical to provide no feedback on participants’ errors. Table 7 presents student participants’ biographic data.
Table 7. Student Participants Biographic Data

<table>
<thead>
<tr>
<th>Demographics / Conditions</th>
<th>Direct WCF</th>
<th>Indirect + WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Semester</strong></td>
<td>2019 Spring, 2019 Summer</td>
<td>2019 Spring, 2019 Summer</td>
</tr>
<tr>
<td>N</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Female</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>Mean age</td>
<td>38</td>
<td>41</td>
</tr>
<tr>
<td>Proficiency level (ACTFL)</td>
<td>Novice</td>
<td>Novice</td>
</tr>
<tr>
<td>L1 English</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Multilingual (more than two languages)</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>Previously visited in China</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Plan to move to China</td>
<td>12</td>
<td>10</td>
</tr>
</tbody>
</table>

A survey about the exposure of the target items was also filled out by the participants at the end of the course. Data from five participants were eliminated during analysis for the following reasons: 1) participants did not follow instructions to only rely on course materials during the three data collection phases (i.e. they reported looking up external materials or sought out external instruction on the target items); or 2) participants demonstrated significant prior knowledge by scoring above 50% when producing particle *le* or classifiers in the pre-treatment first draft. A further two participants’ data was eliminated because the questionnaire revealed they lived together and exchanged their views on the study methods and feedback provided on Writing Task 1. By exchanging knowledge and comparing their feedback, there was insufficient control of their condition and the data no longer reflected solely their treatment group.
While there were four teachers in the Chinese distance learning program, only two participated in the study. The participants were drawn from those students assigned to the two participating teachers. These two teachers were each in charge of delivering individual one-on-one mentoring sessions to a fixed subset of the participants. The WCF treatments were provided by these two teachers: the author of the current study and another teacher “Mary” (pseudonym). Each teacher was in charge of half of the participants. These participants were randomly assigned to one of the two WCF treatment groups, with both teachers having participants in the two treatment categories. Two additional teachers who participated in the design of the research writing tasks and indirect metalinguistic explanations sheet also attended the focus group discussion. All teachers were native speakers of Mandarin Chinese and were bilingual in English with more than 8 years of language teaching experience at the time of the study. All had been teaching the synchronous computer-mediated communication course for the preceding two years at the time of the study, and were familiar with the course content, materials, and online systems. All teachers received instructions and protocols on how to administer the WCF treatment during four one-hour group training discussions. The online training sessions included a lecture describing direct WCF and indirect WCF with metalinguistic explanations, as well as hands-on activities that practiced giving the two types of WCF on the target items to ensure common application and consistency.
3.2 Context

The course that hosted this research study was conducted completely online in an e-learning platform designed and maintained by the hosting institution. The participants navigate through the course independently and are asked to spend between 6 and 8 hours per week studying the online course content (static text, interactive audio, and video). While the online course is only a few years old, the hosting institution has a long history of foreign language instruction.

![Figure 4. Online Course Material for Individual Study](image)

Figure 4. Online Course Material for Individual Study

Figure 4 shows a sample of the online course content that students review individually with multimedia content and an accompanying text. Each week, the participants spend one hour in a teacher-led session on the week’s assigned content. These teacher-led sessions use a web conferencing software that includes a video window, a text box, and a shared workspace for real-time collaboration. See Figure 5 for a screenshot of the software, showing the teacher and
participant (obscured) in the upper left, and the shared workspace at the center. The lower left panel enables the teacher to provide written feedback in an online chat interface with the student.

Over fourteen weeks, the beginner-level course moved through four units with one to four lessons per unit. The course was organized as a task-based language teaching (TBLT) curriculum, built upon a comprehensive needs analysis, and centered around tasks that the participants are likely to encounter when they travel to China in the future. The course schedule showing the organizing tasks for the curriculum are shown below in Figure 6.

Figure 5. Adobe Connect Web Conferencing Software
Lesson objectives included understanding pinyin and characters, introductions, telling date and time, ordering food and drink, and buying household goods. Each unit included descriptions, tasks, and activities for participants to work through at their own pace. These sessions were further refined based on individual needs analysis conducted by the teacher to enable some tailoring of the curriculum to meet participant goals. As an example of a target task, professional introductions may be a focus within the domain of self-introductions. Participants were asked to come to the synchronous computer-mediated communication sessions prepared to review content they studied previously that week in the online modules. Participants were told that the emphasis of the course is the self-guided online modules and course content. They were told that the purpose of the weekly sessions with the teacher was to provide authentic task interactions, to guide progress, and to provide help as needed. The teacher verifies that learners completed at least 80% of the task-based learning objectives in order to pass the course.
3.3 Target Linguistic Items

The complexity and learnability of linguistic structures is interesting when evaluating how researchers have investigated the effectiveness of WCF. Ample evidence in second language acquisition research demonstrates that linguistic structures vary in complexity and have different learning outcomes, and instruction and feedback play an important role (e.g., Sheen, 2007; Li, 2014; Shintani et al., 2014).

Many researchers have made efforts to untangle the relationship between linguistic complexity and types of WCF. Notably, Ferris (1999) first pointed out that “no single form of correction can be effective for all kinds of errors found” (p. 5) and argued that indirect WCF would be most effective where the errors are “patterned and rule-governed.” Many subsequent studies have found supporting evidence for Ferris (1999)’s claim (e.g., Sheen, 2007; Shintani et al., 2014, Shintani & Ellis, 2015). For example, Bitchener et al. (2010) suggested indirect WCF improves accuracy on subject-verb agreement more than direct feedback. Shintani and Ellis (2013) found that indirect WCF was more effective than direct WCF on English articles, while finding both indirect and direct WCF were ineffective on the hypothetical conditional, despite it being a complex rules-based grammatical target. However, Benson and Dekeyser (2018) found a contrasting result where direct WCF was better than indirect on English simple past tense and neither direct nor indirect WCF worked on the present perfect tense on a new writing task. More linguistic structures should be compared in this line of research in order to explore what types of feedback are most effective for different grammatical targets.

Studies of WCF’s effectiveness has been concentrated in only a few contexts and linguistic error domains, limiting the generalizability of findings. More than 83% of WCF studies published from 1992 to 2016 focused on English instruction, with Spanish making up a
sizable part of the remainder (Riazi, Shi & Haggerty, 2018). To better understand the effectiveness of different types of WCF and how to apply it, the research community will benefit from diversifying the sampling of languages examined. There are only sparse examples of less commonly taught languages and no studies examining WCF’s effectiveness in Chinese. This study seeks to contribute to diversifying the knowledge base through its targeting of Mandarin and some of its notable common grammatical structures. Chinese presents unique challenges to native English speakers (Ehrman, 1996; Hu, 2010) that may affect WCF. The grammar, suprasegmental features (i.e., tones), and logographic script are distant from English, with much pedagogical research still focused on fundamental questions about how best to teach these foreign concepts (Bryfonski & Ma, 2019; Hu, 2010; Huang, 2000; Ke, 1996; 2005; Li & Thompson, 1981).

In order to avoid drawing excessive attention to specific target linguistic items in addition to ecological validity issues, this study takes an unfocused approach to WCF and provides students with feedback on all errors. However, in order to allow for consistent measurement of the effects of WCF on written accuracy, the study focuses on two common and basic structures in Chinese that are guaranteed to appear frequently in student’s writing: le and classifiers. Focusing on these targets makes it possible to probe the effects of different feedback types on structures with different linguistic attributes—le, a rules-based target, and classifiers, a more rote target.

The concept of perfective le, drawing on the Aspect Hypothesis (Li & Shirai, 2000) and Vendler’s (1957) work on verb types, encodes an event in its entirety. The particle le is often applied in bounded situations with telic verbs such as verbs encoding accomplishments. Regarding states and activity verbs that encode atelic situations (situations without an endpoint),
an external device (quantifiers) should be added after *le* to indicate a boundary for the action or event. Otherwise, *le* indicates a change of state or new situation. Academic opinion is divided over the linguistic interpretations of *le*. Some researchers hold that there is only one usage of *le*, which marks either termination or completion (e.g., Yang, 2003); other studies and many practitioners maintain that there are two distinct usages for *le*: 1) a perfective *le* that encodes perfectivity and; 2) a change-of-state *le* that indicates inchoativity (e.g., Xiao & McEnery, 2004). The current study follows the two *le* perspective. Examples of perfective *le* and change-of-state *le* are provided below.

**Perfective *le***

**Chinese:** 我吃了饭去学校。

**Pinyin romanization:** Wǒ chī *le* fàn qù xuéxiào.

**Literal translation:** I / eat / particle *le* / meals / go / school

**Translation:** After eating, I went to school. / I will go to school.

**Change-of-state *le***

**Chinese:** 她学第三课了。

**Pinyin romanization:** Tā xué dì-sān kè *le*.

**Literal translation:** She / study / third / lesson / particle *le*

**Translation:** She is on lesson three now.

A classifier in Mandarin is used between a determiner and a noun, usually to combine a quantifier and a noun (e.g., Zhang, 2007). There are semantic and syntactic relations between a classifier and a noun, and the associated noun determines the appropriate classifier.

Semantically, a classifier is needed to categorize objects that share similar inherent properties across a variety of categories. Syntactically, “classifiers are units of enumeration employed to mark countability; their occurrence makes the semantic partitioning of nouns visible” (Wu & Bodomo, 2009, p. 490). Mandarin Chinese is a classifier language and includes more than 900 classifiers (e.g., Zhang, 2007). Learning the appropriate classifier is one of the most challenging
features of Chinese noun phrases for native speakers of English due to the diversity and abstract categorization schema. Classifier rules can seem arbitrary and unrelated for learners, especially when groupings may derive from Chinese historical or cultural context (for example, a dog is classified as a long, thin object). Many Chinese teachers comment that the most practical way for L2 Chinese students to learn the classifier system is rote memorization (e.g., Li & Thompson, 1981; Jiang, 2012). Three examples of classifiers in sentences are provided below:

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Chinese:</th>
<th>Pinyin romanization:</th>
<th>Literal translation:</th>
<th>Translation:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>一条裤子</td>
<td>Yī tiáo kùzi</td>
<td>One / classifier / pants</td>
<td>One pair of pants (tiao is used for long objects).</td>
</tr>
<tr>
<td></td>
<td>一条狗</td>
<td>Yī tiáo gǒu</td>
<td>One / classifier / dog</td>
<td>One dog (tiao is used for long objects).</td>
</tr>
<tr>
<td></td>
<td>两只鸡</td>
<td>Liǎng zhī jī</td>
<td>Two / classifier / chicken</td>
<td>Two chickens (zhi is used for animals)</td>
</tr>
</tbody>
</table>

Based upon previous schemas for structural difficulties (e.g., Ellis, 2007; Goldschneider & DeKeyser, 2005), le and classifiers can be compared in the following dimensions:

1. Perceptual saliency. The learner’s perceived difficulty of a given structure affects how it is acquired. The particle le is always affixed to a verb in the post-verbal position or at the end of a statement. The location of le is opaque. However, a classifier usually precedes the noun. In addition, le is pronounced in a neutral tone while classifiers’ tone is different for each classifier. Beginners usually write in Chinese pinyin and tones since they have
yet to learn Chinese characters. Therefore, classifiers are generally more perceptually salient than the *le*. (Li & Shirai, 2000).

2. Form–meaning mapping. The form–meaning mapping of *le* is opaque because *le* has two different positions in a sentence and the two positions have different interpretations. The form–meaning mapping of classifiers is transparent because a classifier is often used with one category of objects.

3. Explicit knowledge. The rule explanation about usages of *le* is complicated because *le* involves at least three components: (a) the event is finished, (b) the situation must be bounded, and C) the situation is changed. The rule for classifiers is straightforward since a certain classifier is only used with a particular noun.

By targeting *le* and classifiers, two different structures with varying features, the current study sought to probe if different feedback types have differential effects on structures with different linguistic attributes. Teachers’

*3.4 Procedure and Materials*

The study’s activities were spread across a fourteen-week semester. Each activity is displayed in the summary timeline in Figure 7, and further elaborated below.
3.4.1 Background Questionnaire and Think-aloud Protocol Training

All participants provided information about their backgrounds via a questionnaire administered in the first week. According to the results of the background questionnaire, participants were sorted into strata reflective of their gender and language learning background. Prior to treatment, the participants were divided randomly into two feedback groups that received either direct or indirect with metalinguistic explanations on their compositions. Each teacher was

Figure 7. Procedure Timeline Overview
in charge of providing individual instruction to half of the 45 participants according to the assigned treatment.

3.4.2 Writing Tasks

There were two writing tasks, adapted from Li (2014), which were deemed appropriate for the curriculum of the hosting course and were designed to elicit output on the grammar forms targeted in this study. The tasks required the participants to create an original text in response to a situational prompt. Given that the course encourages, but does not require, writing Chinese characters, participants are permitted to write in pinyin romanization, including tone marks if they so choose. Responses are asked to be about 250 words in length. Contextual instructions in each task ensure that the participants output will be roughly comparable.

Writing Task 1

In Writing Task 1, participants were asked to write a narrative story according to the provided scenario. They had one week to write it and reported the amount of time spent completing the task. Within one week after submitting Writing Task 1, participants received WCF according to their group assignment. Revising the first submission served as the first post-test and was submitted two weeks after the first draft. Participants were asked to complete the think-aloud protocol as they reviewed the WCF and completed the Revision of Writing Task 1.

New Writing Task 2

In the eighth and ninth weeks of the course, participants were asked to complete a new piece of writing which was a similar scenario-based writing task with a different context (New Writing Task 2) to address participants’ ability to apply the learning from the Writing Task 1 WCF in a new piece of writing.
Even though all participants finished all writing tasks at home, dictionaries and other tools were not used during writing. Participants had a choice to type and email their assignments as a document, or they could handwrite their assignments and submit a scan or picture of the document. Some participants found typing easier to manage given their study conditions and so the teachers accepted use of the informal tone markings indicated through a number reflecting the formal accent marks used in pinyin. All participants reported writing time and word number when they submitted drafts. The average writing time was 65 minutes, and average word count was 203 words.
3.4.3 Type of Written Corrective Feedback

The feedback treatment that participants received takes the form of either direct WCF or indirect WCF with metalinguistic explanations. Examples of how this feedback appears on the target items are provided below to illustrate the concepts. Mandarin characters are represented in pinyin romanization.

Direct feedback was provided by crossing out the error and inserting explicit corrections. For example, a participant made a lexical error on the Chinese classifier for cars.

Chinese: 我买了一个辆车。

Pinyin romanization: Wǒ mǎi yi gé liàng chē.

Literal translation: I / buy / Particle le / one classifier (general) / classifier (for cars) / car.

Translation: I bought a piece classifier (cars) car.

Indirect with metalinguistic explanations did not provide explicit corrections but rather pointed out the error and provided metalinguistic explanations on how to correct the error. This type of explanation was provided on a separate sheet with the meanings of the encoded indirect feedback, as shown below on the line for metalinguistic feedback.

Chinese: 我去 P-le1 超市。

Pinyin romanization: Wǒ qù P-le1 chāoshì.

Literal translation: I / go / P-le1 / supermarket.
Translation: I go to the supermarket and verb “go” has the error code “P-le1”.

Metalinguistic Feedback: P-le1 means particle le usage 1. Particle le is used to express action completion, such as Wǒ chī le fàn (I ate dinner).

Below are some examples of participants’ feedback from the current study. For both handwritten and typed submissions, the teachers’ hand-wrote feedback on a print-out using a green pen and sent a scan back to the participant for them to review during the think-aloud sessions. In Figure 9, the teacher provided unfocused direct feedback on a participant’s written production of L2 Mandarin. The teacher corrected tone markings, struck out extra or incorrect words, and provided correct forms. For the figure seen in this study, the target grammatical items have added red circles and arrows for emphasis. (Participants were not provided red markings on the target.)

Figure 9. Participant’s error and unfocused direct WCF (in green)
Unfocused direct WCF was provided by crossing out the error and inserting explicit corrections. In Figure 9 above, the generic classifier ge was crossed out and the correct classifier ben was added next to the error. In the next example, le was inserted next to the verb mai (to buy).

In Figure 10, the teacher provided unfocused indirect with metalinguistic explanations to a participant’s written production of L2 Mandarin. Unfocused indirect with metalinguistic explanations appeared in the form of underlining the error with clues or error codes, alongside an accompanying key or sheet (full code sheet in Table 8, below) of explicit rule explanations and examples for the participant to interpret. The abbreviated two-letter codes corresponded to the metalinguistic feedback sheet. The underline told the participant where the code applied. For the purposes of this study, the WCF on the target grammatical items were highlighted in red for emphasis.

Figure 10. Participant’s error and unfocused indirect WCF (with codes and examples in sheet)
Table 8. *Indirect WCF with Metalinguistic Explanation Sheet*  

(Instructions: Feedback on the essays use the abbreviations below. Please reference this sheet to interpret them.)

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Explanation</th>
<th>Example (English meaning, error and correction)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PY</td>
<td>Please check the spelling of the Pinyin</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Error: huǒ PY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct: hao3 / Hǎo</td>
</tr>
<tr>
<td>T</td>
<td>Please check the tone.</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Error: Wo T</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct: wo3 / wǒ</td>
</tr>
<tr>
<td>WW</td>
<td>Wrong Word</td>
<td>I’d like to buy apples.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Error: Wǒ  xiǎhuān WW mái píngguǒ.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct: Wǒ xiǎng mái píngguǒ</td>
</tr>
<tr>
<td>WO</td>
<td>Word Order</td>
<td>I work at FSI.</td>
</tr>
<tr>
<td></td>
<td>Sbj. + Prep + Verb + Obj.</td>
<td>• Error: Wǒ  gōngzuò zài FSI. WO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct: Wǒ zài FSI gōngzuò.</td>
</tr>
<tr>
<td>P-le1</td>
<td>Particle <em>le</em> usage 1</td>
<td>I went to FSI</td>
</tr>
<tr>
<td></td>
<td>Expressing completion with &quot;le&quot;</td>
<td>• Error: Wǒ  qù FSI. P-le1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct: Wǒ qù <em>le</em> FSI.</td>
</tr>
<tr>
<td></td>
<td>He introduced me.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Error: Tā jièshào wǒ. P-le1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct: Tā jièshào <em>le</em> wǒ.</td>
</tr>
<tr>
<td>P-le2</td>
<td>Particle <em>le</em> usage 2</td>
<td>I’m ready.</td>
</tr>
<tr>
<td></td>
<td>Expressing change of state or new situation with &quot;le&quot;</td>
<td>• Error: Hǎo de. P-le2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct: Hǎo <em>le</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I’m in the FSI now.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Error: Wǒ zài <em>le</em> FSI. P-le2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct: Wǒ zài FSI <em>le</em>.</td>
</tr>
<tr>
<td>P-guo</td>
<td>The aspect particle “guó” is used to indicate that an action has been experienced in the past (have been to)</td>
<td>I’ve been to the US.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Error: Wǒ qù Měiguó. Guo</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Correct: Wǒ qù guó Měiguó.</td>
</tr>
<tr>
<td>MW</td>
<td>Measure Word / classifier to indicate quantity of a noun</td>
<td>Examples:</td>
</tr>
<tr>
<td></td>
<td>Number + MW + noun</td>
<td>1 people = Yī  gè rén (the general measure word)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 ducks= Liǎng  zhī yā (for animals)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 boxes of sugar = Sān  hé táng (for box)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 bottles of wine = Sì  píng jiǔ (for bottle)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 novels = Wǔ  běn novels (for books)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 cucumbers= Liù  gèn huángguā (for long and round things)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 ties = Qī  tiáo lingdài (for long and flat things)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8 pairs of socks = Bā  shuāng wàzi (for pair)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9 plates of dishes = Jiǔ  pán cài (for plate )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 bags of rice = Shí  dài dāmí (for bag)</td>
</tr>
</tbody>
</table>
| De | Attributive “de” is used to mark possession or modification. Modifier + de + noun | Examples:  
My dad’s name Wǒ bàba de míngzi  
The apple that I bought Wǒ mǎile de píngguǒ |
|---|---|---|
| RC | Resultative complement is used to describe the result of a verb. Verb + complement | Examples:  
Complete by buying (finished buying) Mǎi hǎo  
Complete by saying (finished saying) Shuō hǎo |
| NOT | Wrong negator used “mei” VS “bu” | Examples:  
I didn’t go FSI. Wǒ méi qù FSI.  
I don’t want to buy hot peppers. Wǒ bù xiǎng mǎi làjiāo |

The code sheet covered eleven separate items and provided a short explanation of the correction alongside its meaning, an example error that showed the usage of the code, and a correction that applied the appropriate correction to the example error. The code sheet included the green colors to convey the feedback that would be written in by the teacher using green pen. Some simple items, like an error in pinyin spelling, are addressed with only one example.

3.4.4 Think-Aloud Protocols

At the outset of the study, participants received a ten-minute training on the think-aloud protocol (see Figure 11 below), including an example audio recording showing notional reactions to feedback and the thought process in writing in Mandarin Chinese.
**Think aloud protocols training**

What is a think aloud?

Think-aloud protocols involve learners voicing their thoughts aloud as they are performing a set of specified tasks. Learners are asked to say aloud whatever comes into their mind as they complete the task. This might include what they are seeing, doing, and feeling.

Why do we think aloud?

Think-aloud data gives teachers and researchers insights into the learner's cognitive processes while completing the task (rather than only the final product), helping support more targeted instruction for students and improvement in course materials.

![How to think aloud?](image)

**Figure 11. Think-Aloud Protocol Training Material**

The think-aloud protocols in the current study were conducted with the teachers present to ensure that they consistently engaged in the protocol and did not fall silent. Prior to beginning the think aloud in the session, the teachers provided a review of the think aloud instructions and reminded the participant that the teacher would not interrupt them until they reached the end of the feedback, and the teacher would not take questions during the exercise. The results from the think aloud varied in length, with some participants reviewing their work in five minutes while
others took as long as twenty minutes. The 48 responses were generally substantive and in-line with the study’s expectations for the target structures.

3.4.5 Feedback Questionnaire and Interview

In the final weeks of the course, all participants were asked to complete a fifteen-item multiple-choice questionnaire asking them about their attitudes toward the feedback they received. (See Appendix H for a copy of the questionnaire.) Participants were also showed examples of direct and indirect feedback and asked to provide their thoughts on the feedback type. In order to ensure that the experimental conditions were effectively controlled, participants were also asked to report any other exposure to the target linguistic items besides the WCF treatment and course materials (e.g., supplementary self-instruction). All of the participants completed the questionnaire, and as discussed in the participants section, some data were excluded from the current study’s analysis due to external additional instruction on the target structure outside of the study, as reported in the questionnaire.

The researcher invited a group of eight participants to participate in voluntary individual interviews, choosing four participants randomly from each experimental group. The selected participants all accepted the invitation and participated in a 20-minute semi-structured interview in the last three weeks of the semester. To assist with their ability to respond to detailed questions about feedback, they were shown examples of direct and indirect feedback with a brief explanation. The interview was audio recorded for subsequent qualitative analysis.

More specifically, the structured interview prompts were provided to participants along with the disclosure (Appendix H) and the topics discussed with participants cover student motivations and the course’s efficacy in meeting those goals; their perceptions and preferences
of the course and assignments; and a series of questions probing their perceptions of the feedback, their engagement with it, and its impact on their learning of the target linguistic items. The researcher used these topics to elicit responses that were further explored in follow-up questions. The interviews were in English.

3.4.6 Teacher Focus Group

In the last week of the semester, the researcher engaged in a video-chat focus group with the other three teachers in the course: this group included the one teacher that provided feedback to the participants during data collection and two teachers who taught in another section of the program not participating in the study. The session was facilitated by the prompts listed in Appendix H and was conducted in the teachers’ L1, Mandarin Chinese. The focus group was intended to include a 20-minute semi-structured series of prompts, but developed into a longer form discussion that included free discussion. The conversation covered the teachers’ perceptions of the course; their engagement and feelings toward the feedback treatments; the teachers’ perception of feedback efficacy and their perceptions of student satisfaction; and their reactions to preliminary statistical results from the current study.
3.5 Data Analyses

3.5.1 Coding of the Think-Aloud Protocols

To address how participants processed direct and indirect WCF with metalinguistic explanations, the study adapts Leow’s (2015) coding scheme to code for levels of depth of processing in the think-aloud data. The modified and more elaborated codes (Caras, 2019) represent depth of processing of the target dichotomies, as prompted by the WCF. Table 9 displays the adapted coding scheme. Depth of processing was only coded when the participant made an error on the target form and on which he/she subsequently received WCF.

Table 9. Depth of Processing Coding Scheme

<table>
<thead>
<tr>
<th></th>
<th>Level 1: Low Depth of Processing</th>
<th>Level 2: Medium Depth of Processing</th>
<th>Level 3: High Depth of Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Shows no potential for processing target</td>
<td>Comments on target in relation to meaning</td>
<td>Arrives at an inaccurate, partially or fully accurate underlying rule</td>
</tr>
<tr>
<td>Descriptors</td>
<td>Reads feedback quickly</td>
<td>Spends more time processing target</td>
<td>Makes hypotheses regarding target</td>
</tr>
<tr>
<td></td>
<td>Translates the phrase to English, says what s/he means in English</td>
<td>Makes comments that indicate some processing of target, refers to a previous instance with the same target</td>
<td>Provides an inaccurate, accurate and/or partially accurate rule</td>
</tr>
<tr>
<td></td>
<td>Repeats target or her/his production in the composition</td>
<td>Some level of cognitive effort to process target</td>
<td>Corrects previous translation</td>
</tr>
<tr>
<td></td>
<td>Says s/he isn’t sure what is wrong, comments on the target but does not indicate any further processing</td>
<td></td>
<td>Spends much time processing target</td>
</tr>
<tr>
<td></td>
<td>Avoids the correction</td>
<td></td>
<td>High level of cognitive effort to process target</td>
</tr>
<tr>
<td></td>
<td>Does not spend much time processing target</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Low level of cognitive effort to process target

Examples from the data

“Emm, Wǒmen chīle yā (we ate ducks). Wǒmen kànle shātān (We saw the beach)...”

“So this part is pretty simple, I don’t have the tone marks. But I know the tones. Next part, Nǐ zuò le shénme (what did you do)? So... I have other words there, but I don’t know.”

“The next sentence is Nana mǎi le. Oh it’s just like the previous sentence showing past tense. So mǎi le”.

“Wǒ lǚ xíng le (I travelled) which means I travelled. So I would say the same thing Wǒ yě lǚ xíng le (I also travelled). It means I also like to travel. Umm... Let’s look at the next one.”

“The next sentence is Nana mǎi le. Oh it’s just like the previous sentence showing past tense. So mǎi le”.

“Wǒ zài Āgēntíng zhù le (I live in Argentina). Emm..I lived in Argentina? Wǒ zhù zài Āgēntíng yī nián le (I live in Argentina for one year now). I lived there for one year? Le? Wǒ gēg mǎi le chē, hěn hǎo de chē (My brother just bought a car, a very nice car). My brother bought a nice car. Le again...past event...So... you just put le in the sentence to mark the past tense? Wǒ yǒu dìtiě. (I use metro). So...

“So the first error that you have le is in the sentence ‘Rènshi...’ I guess Nana hěn gāoxìng (very happy). That type of error on your feedback sheet says particle le usage expressing completion of action. You gave examples and how to correct it...so event in the past, or the English past tense. I wrote, ‘It was nice to know Nana.’ So it would be ‘Rènshile hěn gāoxìng’ (It was nice to meet Nana). Is that correct? Next one...”

Translations of Chinese are in the parenthesis.

The current study applied Zamora’s (2017) depth of processing scoring system to score participants’ think-aloud responses on the target linguistic items. The depth of processing score was calculated by evaluating each think-aloud instance. Low level of processing was given a score of one, instances of medium level were given a score of three, and instances of high level
were given a score of five. For example, if the participant processed all his errors of the target structure at a low level by only reading the WCF, his final depth of processing score was one. However, if the participant processed even one error of the target structures at a high level of processing, especially reporting a partial or accurate underlying rule, and the rest was at a low level, then his final score was 5.

To calculate group average depth of processing score, each participant’s depth of processing score was added then divided by the group participant number 19. The following figure (Zamora 2017) was adopted to interpret each WCF group’s level of depth of processing.

<table>
<thead>
<tr>
<th>Average DOP Score</th>
<th>Corresponding Level of DOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00-2.00</td>
<td>Low</td>
</tr>
<tr>
<td>2.50-3.50</td>
<td>Medium</td>
</tr>
<tr>
<td>4.00-5.00</td>
<td>High</td>
</tr>
</tbody>
</table>

Figure 12. *Depth of Processing Score Ranges (Zamora, 2017)*

The researcher and a colleague independently coded a random sample of 25% of the think-aloud recordings. Interrater agreement was calculated by SPSS (IBM). The inter-rater reliability came to 0.741 (Cohen’s Kappa). The raters discussed their discrepancies (16%) and came to 100% agreement. The researcher then coded the remaining 75% of the recordings. The researcher compared her own provision of WCF and the other rater’s provision by marking a random sample of 25% of the compositions (e.g., Polio & Shea, 2014). The inter-rater agreement came to 100%.
3.5.2 Student Engagement

The study adopted Ellis’s (2010) conceptual framework for student engagement with teacher’s WCF, which included behavioral, cognitive and affective engagement. Analysis of the 38 participants’ written accuracy on particle le and classifiers from the Writing Task 1 to the Revision of Writing Task 1, and New Writing Task 2 was used as an indication of participants’ behavioral engagement. Analysis of the 38 participants’ depth of processing in the WCF think-aloud session was used as an indication of participants’ cognitive engagement. Lastly, 38 participants also filled out questionnaires and attended semi-structured interviews comparing indirect and direct WCF online. They were asked about which type of WCF they preferred; which type made it easier for them to correct the errors; and which type was more beneficial for their language learning and writing in the future. The WCF survey data was used as indication of the participants’ affective engagement. Qualitative analysis was used to inform and provide additional information on the quantitative findings. The feedback interviews, and open-ended survey item responses, as well as teacher field notes, were coded using a thematic, grounded approach (following Mackey & Gass, 2012). The resulting themes were then associated with relevant excerpts from the audio-recordings of participants’ performances.

3.5.3 Scoring

The written output of the target linguistic items from the baseline Writing Task 1, Revision and New Writing Task 2 for both experimental groups were scored for accuracy. Each writing task elicited usages of le and classifiers given the length of the assignment and their frequency of use. The accuracy score calculation was made using the formula for Target-Like Use Analysis applied in Pica (1983). In Target-like Use Analysis, morphemes are first scored in obligatory contexts. Then the score is divided by the sum of both obligatory contexts and non-
obligatory contexts (where the form is produced unnecessarily or inappropriately). The formula for Target-Like Use Analysis is presented in Figure 13 below.

\[
TLU = \frac{n \text{ correct suppliance in obligatory contexts}}{|n \text{ obligatory contexts}| + |n \text{ suppliance in non-obligatory contexts}|}
\]

Figure 13. Formula for Target-Like Use Analysis of Morphemes from Pica (1983)

3.5.4 Statistical Analyses

For each of the two treatment groups, the accuracy scores obtained in Writing Task 1, the Revision of Task 1, and New Writing Task 2 were submitted to a 2 x 3 repeated-measures ANOVA after confirming the assumption of normality and homogeneity of the data. Type of WCF (direct or indirect with metalinguistic explanations) was entered as a between-subjects factor, and time (Writing Task 1, Revision of Writing Task 1, and New Writing Task 2) was entered as a within-subjects factor. Significance testing and effect sizes (Cohen’s \(d\)) were also calculated using SPSS. Sphericity was checked to ensure the data met the assumptions of repeated-measures ANOVA. The sphericity condition applies where the variances of the differences between all possible pairs of within-subject conditions are equal (Mauchly, 1940).

Also, the current study followed Plonsky & Oswald’s (2014) field-specific range for within groups Cohen’s \(d\), which is as follows: 0.60 is small, 1.00 is medium, and 1.40 is large; between groups Cohen’s \(d\) which is as follows: 0.40 is small, 0.70 is medium, and 1.00 is large.
CHAPTER IV: RESULTS

This chapter presents quantitative and qualitative results to answer the four research questions. It follows Larson-Hall’s (2015) guidelines for reporting quantitative data. This section reports data from 38 participants, as discussed in the preceding chapter.

4.1 Research Question 1

Research Question 1 asked “What is the effect of different types of WCF (direct vs. indirect with metalinguistic explanations) on low-proficiency L2 Chinese writers’ accuracy with le and classifiers?”

4.1.1 Statistical Analysis of Accuracy Results for Particle Le

Table 10 displays the mean scores for le from all three tasks. In both treatments, the participants started with very low accuracy in Writing Task 1, with mean accuracy of 9% and 7%—with many having no correct usages in their first production. Both groups improved markedly after feedback, correcting the vast majority of their mistakes following feedback, achieving 98% and 88% accuracy. However, in New Writing Task 2, the performance gains appeared to persist for indirect WCF with metalinguistic explanation, achieving 78% mean accuracy. The direct group’s gains decreased substantially to a 12% mean. This difference in accuracy rates is displayed in Figure 14.
Table 10. Descriptive Statistics for Particle Le Accuracy Scores by Type of WCF

<table>
<thead>
<tr>
<th>Type of WCF</th>
<th>Writing Task 1</th>
<th>Revision of Writing Task 1</th>
<th>New Writing Task 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Direct</td>
<td>19</td>
<td>9% (0.09)</td>
<td>98% (0.05)</td>
</tr>
<tr>
<td>Indirect+</td>
<td>19</td>
<td>7% (0.10)</td>
<td>88% (0.13)</td>
</tr>
</tbody>
</table>

“Indirect +” means indirect WCF with metalinguistic explanations

This difference in accuracy rates was statistically tested using an ANOVA as shown in the results tables below and was visualized in Figure 15. Before reporting the within groups statistics, sphericity was checked to ensure the data met the assumptions of repeated-measures ANOVA. For this study, Mauchly’s Test of Sphericity was statistically significant (p = 0.004, Greenhouse-Geisser = 0.78, Huynh-Feldt = 0.84), showing a violation of sphericity occurred and the variances of the differences between all combinations of the conditions were not equal.
Girden’s recommended solution to this situation is to apply a corrective factor to the results. When Greenhouse-Geisser is > 0.75, the Huynh–Feldt correction should be applied and when epsilon is < 0.75 or nothing is known about sphericity, the Greenhouse–Geisser correction should be applied (Girden, 1992). Therefore, this study reports Huynh–Feldt statistics in le comparisons.

The repeated measures ANOVA revealed significant main effects for Type of WCF (between-subjects ANOVA $F (1, 35) = 46.34, p = 0.00$) and Time ($F (1.67, 58.45) = 496.59, p = 0.00$), and a significant interaction between Type of WCF and Time ($F (1.67, 58.45) = 118.72, p = 0.00$) as reported in Table 11 and 12. This confirmed that the accuracy results in each phase of the study were different. The results from the interaction between Time and Groups showed that the WCF treatment group also mattered in each phase. The main effect found for Type of WCF should be interpreted in light of its significant interaction with Time, visually displayed in Figure 15, below.

Table 11. Summary of Between-Subject Analysis of Variance for Particle Le

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>26.29</td>
<td>1345.04</td>
<td>0.00</td>
<td>0.98</td>
</tr>
<tr>
<td>WCF Group</td>
<td>1</td>
<td>0.91</td>
<td>46.34</td>
<td>0.00</td>
<td>0.57</td>
</tr>
<tr>
<td>Error</td>
<td>35</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $p$ is significant at ≤ 0.05.

Table 12. Summary of Within-Subject ANOVA of Accuracy Scores

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1.67</td>
<td>8.01</td>
<td>496.59</td>
<td>0.00</td>
<td>0.93</td>
</tr>
<tr>
<td>Time*Group</td>
<td>1.67</td>
<td>1.19</td>
<td>118.72</td>
<td>0.00</td>
<td>0.77</td>
</tr>
<tr>
<td>Error</td>
<td>58.45</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $p$ is significant at ≤ 0.05.
The effect size between Groups at each time are reported in Table 14 below. There was a small effect size in Writing Task 1, with $d = 0.21$. In revision, direct feedback had a medium effect size of 1.02. Most notably, the effect size in New Writing Task 2 was 3.82, a large effect size.

In order to further explore the differences in outcomes between groups, a post-hoc pairwise comparison analysis is presented in Table 13. This analysis compares the results from the two experimental groups at each phase. The pairwise comparison analysis found that the two groups’ means are not significantly different in their initial performance. However, significant differences in performances were found between the groups on the Revision of Writing Task 1 and New Writing Task 2. At the time of Revision of Writing Task 1, the direct WCF group
outperformed the indirect group with metalinguistic explanations \((p = 0.003)\). However, on New Writing Task 2 the indirect-with-metalinguistic-explanations group outperformed the direct group \((p < 0.01)\)—the data indicated indirect WCF with metalinguistic explanation resulted in 66% higher mean accuracy.

Table 13. Post-hoc Pairwise Comparison for Between Groups Particle Le Accuracy Scores on Revision of Writing Task 1 and New Writing Task 2

<table>
<thead>
<tr>
<th>Time</th>
<th>WCF Type</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>P</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing Task 1</td>
<td>Direct VS Indirect+</td>
<td>0.02</td>
<td>0.03</td>
<td>0.62</td>
<td>[-0.05, 0.08]</td>
</tr>
<tr>
<td>Revision of Writing Task 1</td>
<td>Direct VS Indirect+</td>
<td>0.10</td>
<td>0.03</td>
<td>0.003</td>
<td>[0.04, 0.17]</td>
</tr>
<tr>
<td>New Writing Task 2</td>
<td>Direct VS Indirect+</td>
<td>-0.66</td>
<td>0.06</td>
<td>0.00</td>
<td>[-0.77, 0.55]</td>
</tr>
</tbody>
</table>

Note. \(p\) is significant at \(\leq 0.05\). “Indirect +” means indirect WCF with metalinguistic explanations

Table 14. Effect Size for Le Mean Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Writing Task 1</th>
<th>Revision of Writing Task 1</th>
<th>New Writing Task 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect+ VS.</td>
<td>(d = 0.21);</td>
<td>(d = 1.02);</td>
<td>(d = 3.82);</td>
</tr>
<tr>
<td>Direct</td>
<td>Small</td>
<td>Large</td>
<td>Large</td>
</tr>
</tbody>
</table>

“Indirect +” means indirect WCF with metalinguistic explanations

Table 15 Reports the difference in mean accuracy between the time periods for each type of WCF.
Table 15. Post-hoc Pairwise Comparison for Within WCF Groups Particle Le Accuracy Scores on Writing Task 1, Revision of Writing Task 1 and New Writing Task 2

<table>
<thead>
<tr>
<th>Time</th>
<th>WCF Type</th>
<th>Mean Difference (A-B)</th>
<th>Std. Error</th>
<th>P</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing Task 1 VS Revision of Task 1</td>
<td>Direct</td>
<td>-0.89</td>
<td>0.03</td>
<td>0.00</td>
<td>[-0.94, -0.84]</td>
</tr>
<tr>
<td>Revision VS New Writing Task 2</td>
<td>Direct</td>
<td>0.86</td>
<td>.004</td>
<td>0.00</td>
<td>[0.77, 0.94]</td>
</tr>
<tr>
<td>Writing Task 1 VS New Writing Task 2</td>
<td>Direct</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.42</td>
<td>[-0.12, 0.05]</td>
</tr>
<tr>
<td>Writing Task 1 VS Revision of Task 1</td>
<td>Indirect +</td>
<td>-0.81</td>
<td>0.03</td>
<td>0.00</td>
<td>[-0.86, -0.75]</td>
</tr>
<tr>
<td>Revision VS New Writing Task 2</td>
<td>Indirect +</td>
<td>0.09</td>
<td>0.04</td>
<td>0.03</td>
<td>[0.01, 0.17]</td>
</tr>
<tr>
<td>Writing Task 1 VS New Writing Task 2</td>
<td>Indirect +</td>
<td>-0.71</td>
<td>0.04</td>
<td>0.00</td>
<td>[-0.79, -0.62]</td>
</tr>
</tbody>
</table>

Note. $p$ is significant at $\leq 0.05$. “Indirect +” means indirect WCF with metalinguistic explanations.

The analysis for direct WCF shows a statistically significant increase in revision accuracy from Writing Task 1 ($p = 0.00$), a significant decrease in performance from Revision of Writing Task 1 to New Writing Task 2 ($p = 0.00$), but no significant difference between Writing Task 1 and New Writing Task 2 accuracy ($p = 0.42$). In contrast, the indirect WCF results show a similar increase in the revision accuracy from Writing Task 1 ($p = 0.00$), a significant decrease in performance from Revision of Writing Task 1 to New Writing Task 2 ($p = 0.03$), but maintained a significant increase in accuracy on New Writing Task 2 when compared to Writing Task 1.
These results provide strong statistical evidence for discussing the role of Type of WCF in learning outcomes between the two groups.

The effect sizes are calculated using Cohen’s $d$ and are reported in Table 14 for $le$. The direct WCF for $le$ has a small 0.25 positive effect size moving from Writing Task 1 to New Writing Task 2, suggesting the participants accuracy improved little from the feedback treatment. The comparison of Writing Task 1 and New Writing Task 2 accuracy scores shows that indirect WCF with metalinguistic explanations had a large 4.49 effect size. When compared to direct WCF, this suggests that the indirect treatment had a larger positive effect on accuracy scores.

4.1.2 Statistical Analysis of Accuracy Results for Classifiers

This subsection completes the same analysis as above, but for the results on Chinese classifiers. Table 16 displays the mean scores for classifiers from all three tasks, finding a similar pattern in both treatment groups. The participants from both the direct and indirect WCF with metalinguistic explanations groups started with low mean accuracy of 16%, then improved based on feedback in the Revision, but reverted back to low accuracy in New Writing Task 2. There was little difference between the groups and neither group seemed to have had major improvement as a result of the feedback treatment. Similar to $le$ analysis, an ANOVA was completed to find statistically significant differences between these groups.

Table 16. Descriptive Statistics for Classifier Accuracy Scores per Written Corrective Feedback Type

<table>
<thead>
<tr>
<th>WCF Type</th>
<th>Writing Task 1</th>
<th>Revision of Writing Task 1</th>
<th>New Writing Task 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>N 19</td>
<td>M (SD) 16% (0.16)</td>
<td>M (SD) 87% (0.27)</td>
</tr>
<tr>
<td>Indirect +</td>
<td>N 19</td>
<td>M (SD) 16% (0.17)</td>
<td>M (SD) 73% (0.14)</td>
</tr>
</tbody>
</table>

“Indirect +” means indirect WCF with metalinguistic explanations
The Mauchly’s Test of Sphericity is statistically significant \((p < 0.01, \text{Greenhouse-Geisser} = 0.72, \text{Huynh-Feldt} = 0.77, \text{Lower-bound} = 0.50)\). Therefore, violation of sphericity occurred and the variances of the differences between all combinations of the conditions are not equal. Girden’s recommended solution to this situation is to apply a corrective factor to the results. When Greenhouse-Geisser is > 0.75, the Huynh–Feldt correction should be applied and when epsilon is < 0.75 or nothing is known about sphericity, the Greenhouse–Geisser correction should be applied (Girden, 1992). Therefore, this study reports report Greenhouse-Geisser statistics in the classifiers’ comparison.

The repeated measures ANOVA revealed a significant main effect for Time \((F (1.44, 52.05) = 260.14, p = 0.00)\), and a significant interaction between Type of WCF and Time \((F (1.45, 52.05) = 5.15, p = 0.02)\), but not for Type of WCF (between-subjects ANOVA \(F (1, 36) = 0.42 \ p = 0.52\)). The WCF treatment groups were not significantly different from each other.
irrespective of time, as reported in Table 16 and 17. This confirmed that the accuracy results in each phase of the study were similar between two groups. The main effect found for Type of WCF should be interpreted in light of its significant interaction with Time, visually displayed in Figure 17, below.

Table 17. Summary of Between-Subject Analysis of Variance for Classifiers

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>17.47</td>
<td>345.36</td>
<td>0.00</td>
<td>0.91</td>
</tr>
<tr>
<td>WCF Group</td>
<td>1</td>
<td>0.02</td>
<td>0.42</td>
<td>0.52</td>
<td>0.01</td>
</tr>
<tr>
<td>Error</td>
<td>36</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $p$ is significant at $\leq 0.05$.

Table 18. Summary of Within-Subject Analysis of Variance for Classifier Accuracy Scores

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1.44</td>
<td>6.65</td>
<td>260.14</td>
<td>0.00</td>
<td>0.88</td>
</tr>
<tr>
<td>Time*Group</td>
<td>1.45</td>
<td>0.13</td>
<td>5.15</td>
<td>0.02</td>
<td>0.13</td>
</tr>
<tr>
<td>Error</td>
<td>52.05</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $p$ is significant at $\leq 0.05$. 

99
A post-hoc pairwise comparison analysis is presented in Table 18 showing that there was a significant improvement in mean accuracy for both written corrective treatments reflected in the revision, however, there was a significant decrease going into New Writing Task 2, to the point where Writing Task 1 and New Writing Task 2 accuracy was significantly different for the indirect-with-metalinguistic-explanations group and was not significantly different for the direct group. The main effect for time for the direct group was statistically significant ($p = 0.00$) with an overall mean change of -71%. The main effect for time for indirect group was also statistically significant ($p = 0.00$) with an overall mean change of -57%. From Revision of Writing Task 1 to New Writing Task 2, both groups decreased in classifier accuracy score. The main effect for time for the direct group was statistically significant ($p < 0.01$) with an overall
mean change of 86%. The main effect for time for indirect metalinguistic explanation group was statistically significant ($p < .01$) with an overall mean change of 49%. Finally, comparing Writing Task 1 with New Writing Task 2, the main effect for time for the direct group was not statistically significant ($p = 0.31$) with an overall mean change of -3%. It was statistically significant ($p = 0.01$) for indirect WCF with metalinguistic explanation group and the effect size was small (0.51).

Table 19. *Post-hoc Pairwise Comparison for Between Groups Classifier Accuracy Scores on Revision of Writing Task 1 and New Writing Task 2*

<table>
<thead>
<tr>
<th>Time</th>
<th>WCF Type</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>P</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing Task 1</td>
<td>Direct VS. Indirect+</td>
<td>-1%</td>
<td>0.05</td>
<td>0.89</td>
<td>[-0.14, 0.10]</td>
</tr>
<tr>
<td>Revision of Writing Task 1</td>
<td>Direct VS. Indirect+</td>
<td>14%</td>
<td>0.07</td>
<td>0.05</td>
<td>[-0.001, 0.28]</td>
</tr>
<tr>
<td>New Writing Task 2</td>
<td>Direct VS. Indirect+</td>
<td>-5%</td>
<td>0.04</td>
<td>0.21</td>
<td>[-0.13, 0.29]</td>
</tr>
</tbody>
</table>

Note. $p$ is significant at $\leq 0.05$. “Indirect +” means indirect WCF with metalinguistic explanations

Table 20. *Effect Size Between Groups for Classifiers*

<table>
<thead>
<tr>
<th></th>
<th>Task 1</th>
<th>Revision</th>
<th>Task 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect+ VS. Direct</td>
<td>$d = 0$;</td>
<td>$d = 0.65$</td>
<td>$d = 0.55$</td>
</tr>
<tr>
<td>No effect</td>
<td>Small</td>
<td>Small</td>
<td>Small</td>
</tr>
</tbody>
</table>

“Indirect +” means indirect WCF with metalinguistic explanations
Table 21. Post-hoc Pairwise Comparisons for Within WCF Groups for Classifier Scores

<table>
<thead>
<tr>
<th>Time</th>
<th>WCF Type</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>P</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing Task 1 VS</td>
<td>Direct</td>
<td>-0.71</td>
<td>0.05</td>
<td>0.00</td>
<td>[-0.82, -0.61]</td>
</tr>
<tr>
<td>Revision of Task 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revision VS New Writeing</td>
<td>Direct</td>
<td>0.685</td>
<td>0.05</td>
<td>0.00</td>
<td>[0.59, 0.78]</td>
</tr>
<tr>
<td>Task 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing Task 1 VS</td>
<td>Direct</td>
<td>-0.03</td>
<td>0.03</td>
<td>0.31</td>
<td>[-0.09, 0.03]</td>
</tr>
<tr>
<td>New Writing Task 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing Task 1 VS</td>
<td>Indirect+</td>
<td>-0.57</td>
<td>0.53</td>
<td>0.00</td>
<td>[-0.67, -0.46]</td>
</tr>
<tr>
<td>Revision of Task 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Revision VS New Writing</td>
<td>Indirect+</td>
<td>0.49</td>
<td>0.05</td>
<td>0.00</td>
<td>[0.40, 0.60]</td>
</tr>
<tr>
<td>Task 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing Task 1 VS</td>
<td>Indirect+</td>
<td>-0.73</td>
<td>0.03</td>
<td>0.01</td>
<td>[-0.13, -0.16]</td>
</tr>
<tr>
<td>New Writing Task 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. p is significant at ≤ 0.05. “Indirect+” means indirect WCF with metalinguistic explanations.

The effect sizes for classifiers are calculated using Cohen’s $d$. For classifiers, both direct and indirect WCF with metalinguistic explanation exhibited similar scores and patterns. Both treatments had a small positive effect size of 0.07 and 0.51 respectively going from Writing Task 1 to New Writing Task 2, suggesting the feedback had little durable effect on classifiers.

4.1.3 Summary of Research Question One Results

The first research question asked: “What is the effect of different types of WCF (direct vs. indirect with metalinguistic explanation) on low-proficiency L2 Chinese writers’ accuracy with le and classifiers? It was expected that both groups would see improvement in Revision of Writing Task 1 after receiving feedback, followed by some attrition in the delayed post-test New
Writing Task 2. The results revealed differences in how different types of WCF were received for different target items.

On the particle le, the direct WCF group outperformed the indirect-WCF-with-metalinguistic-explanation group on Revision of Writing Task 1, while the indirect-with-metalinguistic-explanations group outperformed the direct group on New Writing Task 2. The between-subjects repeated-measures ANOVA yielded significant main effects for Type of WCF and Time, and a significant interaction between Type of WCF and Time. A post-hoc pairwise comparison analysis revealed significant differences at the time of Revision of Writing Task 1 and New Writing Task 2 for the direct and indirect-with-metalinguistic-explanations groups. Both groups significantly improved after each WCF treatment on the Revision of Writing Task 1. However, while both groups’ accuracy decreased significantly from the Revision of Writing Task 1 to New Writing Task 2, only those in the indirect-with-metalinguistic-explanations group maintained significantly improved accuracy in usage of particle le after two weeks on New Writing Task 2 with a large effect size.

As for classifiers, the pairwise analysis found both groups significantly improved after each WCF treatment on the Revision of Writing Task 1, and significantly decreased in New Writing Task 2. The between-subjects repeated measures ANOVA also showed a significant main effect for Time, and a significant interaction between Type of WCF and Time, but not for Type of WCF. The direct feedback group outperformed the indirect-with-metalinguistic-explanations group on the Revision of Writing Task 1, but no significant difference in accuracy was found on New Writing Task 2. Similar to the le results, only those in the indirect-with-metalinguistic-explanations group maintained a significantly improved accuracy in usage of classifiers after two weeks in New Writing Task 2 with a small effect size.
4.2 Research Question 2

The second research question asked, “How do low-proficiency L2 Chinese writers process the two different types of WCF (direct vs. indirect with metalinguistic explanations) on their written production?” The data in this subsection explains the differences in mean accuracy observed in the data under Research Question 1. According to the coding schema, each participant was given a depth of processing score. If the participant was able to process the target at high DoP at any point in the think aloud, the participant’s final depth of processing score is rated as high. The results are further evaluated through excerpts from the think-aloud protocols to exemplify participants’ depth of processing from the direct and indirect WCF groups.

4.2.1 Depth of Processing for Particle Le

Table 21 shows the percentage of participants, by direct and indirect WCF with metalinguistic explanations group, who processed at each level of DoP for particle le. In the direct group, 15 (78%) participants display a low DoP, with 2 (11%) exhibiting medium and high DoP. In the indirect WCF with metalinguistic explanations group, only 1 (5%) displayed low DoP, with 4 (21%) showing medium DoP. The majority, 14 (74%), exhibited high-DoP. The large difference in level of depth of processing between the groups is visualized in Figure 18.

<table>
<thead>
<tr>
<th>Levels of DoP</th>
<th>Direct WCF Group</th>
<th>Indirect +</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>78%, n = 15</td>
<td>5%, n = 1</td>
</tr>
<tr>
<td>Medium</td>
<td>11%, n = 2</td>
<td>21%, n = 4</td>
</tr>
<tr>
<td>High</td>
<td>11%, n = 2</td>
<td>74%, n = 14</td>
</tr>
</tbody>
</table>

“Indirect +” means indirect WCF with metalinguistic explanations.
**Figure 18: Depth of Processing Results for Particle Le per WCF Type**

Excerpts from participants’ think-aloud responses while completing particle *le* tasks are displayed in Table 22. Chinese pinyin romanization is shown in italics and English translations are provided in parentheses. Some of the metalinguistic WCF participants mentioned were the metalinguistic abbreviations or the metalinguistic explanations provided on the explanations sheet (See Appendix E for the full list). Capitalized words represent participants’ emphasis, ellipsis represent pauses/hesitations, and underlined words indicate errors.

Of the fifteen participants in the direct group that processed particle *le* WCF at a low depth of processing level, most ignored corrections to their application of *le* and did not verbalize any processing of *le*. They all moved on to other salient WCF such as pronunciation, and nouns.
Table 23. Examples of Direct WCF with Low Level of Depth of Processing on Particle Le

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. “So this part is pretty simple, I don’t have the tone marks. But I know the tones. Next part, Nǐ zuò le shénme (what did you do)? So… I have other words there, but I don’t know.”</td>
</tr>
<tr>
<td>2. “The next sentence I try to say is I studied for two years. So Wǒ xué xí le liǎng nián le (I have studied Chinese for two years). Umm… I don’t know this. I need some help on this. Maybe we’ll go back to this later.”</td>
</tr>
<tr>
<td>3. “A lot of English here. Hehe….Wǒ chī le hǎo chī de cài (I ate delicious dishes). Oh, I see that’s nice or delicious food. Chī (to eat) is eat, and hǎo chī (delicious)” that’s good eat, delicious. Wǒ qù le hǎowán de dì fāng (I went to interesting places). Yeah, “qù le hǎowán de dìfāng (went to interesting place). Umm, good!”</td>
</tr>
<tr>
<td>4. “So I’m reading your comments in my ….haha. What’s the thing next to Wǒ? I don’t understand. Wǒ xué le (I studied). That’s with “X”. This one I can’t say it.”</td>
</tr>
</tbody>
</table>

For example, in Excerpt 1, above, the participant did not put any tones on his submitted draft text. When he got the draft with direct WCF adding in the correct tones, he immediately paid attention to the tone marks and ignored the le corrections. Le is pronounced with a neutral tone, and is written in pinyin without a tone mark, so a correction to le usage would be easily ignored by a participant focused solely on the tone markings. In Excerpt 3, the participant also passed by the le correction in “Wǒ chī le hǎo chī de cài (I ate delicious dishes).” Interestingly, even though she reproduced the le in reading the feedback, she focused on the noun phrase after the action “chī de cài (delicious dishes).” In Excerpts 2 and 4, above, the participants noticed the le correction. They can see it has been added, but do not process a reason for why it was added, and were unable to process its usage, so they moved on. In Excerpt 2, the participant even reflected that they did not understand this correction and reflected that they would need to return to this sentence to better understand why.
Table 24. Examples of Direct WCF with Medium-Level of Depth of Processing on Particle Le

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. “Wǒ lǚ xíng le (I travelled) which means I travelled. So I would say the same thing Wǒ yě lǚ xíng le (I also travelled). It means I also like to travel. Umm… Let’s look at the next one.”</td>
</tr>
<tr>
<td>6. “I appreciate the smaller corrections that is the le has been added many times and some of the words that I wrote as tāo méi, that should be a c, Cǎo méi (strawberry). But these make sense to me, Ok! Thank you.”</td>
</tr>
</tbody>
</table>

Table 25. Examples of Direct WCF with High-Level of DoP on Particle Le

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. “Wǒ zài Āgēntíng zhù le (I live in Argentina). Emm… I lived in Argentina? Wǒ zhù zài Āgēntíng yī nián le (I live in Argentina for one year now). I lived there for one year? Le? Wǒ gēge mǎi le chē, hěn hǎo de chē (My brother just bought a car, a very nice car). My brother bought a nice car. Le again…past event…So… you just put le in the sentence to mark the past tense? Wǒ yǒu dìtiě. (I use metro). So…</td>
</tr>
<tr>
<td>8. “Ok, we need to talk about le, so le, umm…, it means happened in the past? So qù le (went), gào su le (told), mǎi le (bought)…So a lot of le-s. Ok! Méi wèn tí (no problem)!”</td>
</tr>
</tbody>
</table>

There were two participants who processed particle le at a medium DoP level. Both noticed the le corrections and spent some mental effort to try to understand the le structure. However, in both cases, the participants did not truly understand what they were being shown. In Excerpt 5, the participant incorrectly translated the phrase “I also traveled,” and in Excerpt 6, the participant only noted the frequency of the le feedback. In the full think-aloud responses these participants did not demonstrate a full understanding of the le feedback, even though they reproduced it in the specific case.
Only two participants in the direct WCF group achieved a high depth of processing level on *le*. Both participants noticed many *le* corrections and spent time analyzing the rule or usage of *le*. Especially, when these two participants finished looking at the WCF in the draft, they noticed *le* was added many times, and that its usage was mostly associated with verbs. In the think-aloud responses for both Excerpt 7 and 8, the participants concluded that *le* should be added after the action to convey the past tense in English.

In contrast, only one participant from the indirect WCF group processed particle *le* at the low level. In Excerpt 9 (below), he noticed the *le* code and took a look at the metalinguistic feedback sheet, but was not able to understand the correction. In the think aloud session, he incorrectly believed that it was a new structure and felt that the rules were too complicated to understand within the short time of a lesson. He moved on to focus on the content that he had prior knowledge of, such as speech order and different ways to express the number two in Chinese.

Table 26. Examples of Low-Level of Depth of Processing on Particle *Le*: Indirect WCF with Metalinguistic Explanations

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. “Nana tái tái mǎi le èr (Nana bought two)… umm… I don’t understand the particle <em>le</em>. Umm…the word choice, here, umm… should be Liǎng (two). So counting objects you don’t use èr (two). You use Liǎng (two)…”</td>
</tr>
<tr>
<td>10. “Emm, Wǒmen chī le yā (we ate ducks). Wǒmen kàn le shātān (We saw the beach)…”</td>
</tr>
</tbody>
</table>
Table 27. Examples of Medium-Level of Depth of Processing on Particle \textit{le}: Indirect WCF with Metalinguistic Explanations

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. “So the next one is \textit{Wǒ xué xì Zhōng wén} (I study Chinese). Then between there is a \textit{le} usage… So \textit{Wǒ xué xì le zhōng wén yī nián} (I studied Chinese for one year)” …. “\textit{Wǒ chī le xiā} (I ate shrimp)…So I should add the \textit{le} for the same reason.”</td>
</tr>
<tr>
<td>13. “The next sentence is \textit{Nana mǎi le}. Oh it’s just like the previous sentence showing past tense. So mǎi \textit{le}”</td>
</tr>
</tbody>
</table>

Four participants processed \textit{le} at the medium depth of processing level. These participants made comments that indicated some processing of the target or referred to a previous instance with the same \textit{le} usage. For example, the participant in Excerpt 11 figured out the \textit{le} correction earlier in the sentence “\textit{Wǒ xué xí le Zhōng wén} (I studied Chinese)” and realized that \textit{le} should be added after the verb to indicate past tense in English, commenting “I should add \textit{le} for the same reason” without checking the metalinguistic feedback sheet. He subsequently accurately processed “\textit{Wǒ chī le xiā} (I ate shrimp)” using the past tense. Similarly, the participant in Excerpt 13 correctly recognized the pattern from a previous correction and voiced the past tense reasoning for using \textit{le}.
Table 28. Examples of High-Level of Depth of Processing on Particle Le: Indirect WCF with Metalinguistic Explanations

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>14. “Wǒ chīle shālā (I ate salad). Adding <em>le</em> because you are showing, describing the whole trip about, umm, the time in past. You need to, umm, you don’t have past tense verbs. You are adding <em>le</em> to indicate this action is in the past.”</td>
</tr>
<tr>
<td>15. “Emm, Wǒ rènshíle nǐ hěn gāoxìng. So it’s the particle usage and because I’m saying it was nice to meet you? Or it’s nice to meet you and that action is finished. So I need to added a <em>le</em> because the action has been completed.”</td>
</tr>
<tr>
<td>16. “So the first error that you have <em>le</em> is in the sentence ‘Rènshi…’ I guess Nana hěn gāoxìng (very happy). That type of error on your feedback sheet says particle <em>le</em> usage expressing completion of action. You gave examples and how to correct it…so event in the past, or the English past tense. I wrote, ‘It was nice to know Nana.’ So it would be ‘Rènshíle hěn gāoxìng’ (It was nice to meet Nana). Is that correct? Next one…”</td>
</tr>
<tr>
<td>17. Umm, “Wǒ xuéxí le Zhōngwén (I studied Chinese).” There is a particle usage two. Umm, the new situation with <em>le</em>. So the error is “Wǒ zài le FSI (I’m at School).” The correct way would be “Wǒ zài School le (I’m at FSI).” So putting it after the object? Umm…So this particular piece would be “Wǒ xuéxíle Zhōngwén yī nián le (I have studied Chinese for one year).” Umm… yī nián le …umm…</td>
</tr>
</tbody>
</table>

There were 14 participants who demonstrated a high depth of processing on *le* in the indirect WCF group. In the excerpts above, they identified the usage of *le* as a past tense marker, such as in Excerpt 14’s statement that it is a substitute for the lack of verb conjugation in Chinese. In Excerpt 15, the participant expressed another interpretation of *le* usage—to mark task completion. In each of the high-level depth of processing responses, there was recognition of the metalinguistic basis for why *le* is applied in the given WCF.
4.2.2 Depth of Processing for Classifiers

Table 28 shows the percentage of participants, by direct and indirect group, who processed at each level of depth of processing for classifiers. The direct group was weighted toward lower depth of processing, with a majority (63%) of participants exhibiting low depth of processing. The indirect WCF with metalinguistic explanations group was evenly distributed across the three levels of depth of processing—low, medium, and high. These distinct patterns in the depth of processing distribution are visualized in Figure 19.

Table 29. Percentage of Participants by Level of Depth of Processing for Classifiers.

<table>
<thead>
<tr>
<th>Levels of DoP</th>
<th>Direct WCF Group</th>
<th>Indirect WCF + Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>63%, n = 12</td>
<td>37%, n = 7</td>
</tr>
<tr>
<td>Medium</td>
<td>32%, n = 6</td>
<td>37%, n = 7</td>
</tr>
<tr>
<td>High</td>
<td>5%, n = 1</td>
<td>26%, n = 5</td>
</tr>
</tbody>
</table>

Figure 19. Depth of Processing Results for Classifiers per WCF Type
The 12 participants who displayed low-levels of depth of processing on classifiers generally had no meaningful reflection on the classifier, beyond some repetition of the correction. In Example 18, the participant repeated the full phrases, including the direct WCF on the number and classifier for the nouns, but they were entirely focused on memorizing the nouns, ignoring the classifiers. Other excerpts show the participants repeated the correct language as if to aid in rote memorization, but then continue on to the next item.

Table 30. **Examples of Direct WCF with Low-Level of Depth of Processing on Classifiers**

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>18. “He yī hé cǎoméi (Box, one box of strawberries), tomatoes? Yī bǎ xiāngjiāo (One bunch of bananas), xiāngjiāo is bananas. I forgot this. Hé wǒ mái le yī zhī ji. So ji is chicken, I don’t know.”</td>
</tr>
<tr>
<td>19. “Wǒ mái le yī gé píngguǒ, liǎng píng jiàngyóu, yī jīn xīlánhuā (….that’s hard. I’m done!”</td>
</tr>
<tr>
<td>21. “Wǒ mái le yī zhī jì, liǎng gēn huángguā, yī gé cǎoméi, yī bǎ xiāngjiāo, sān gè píngguǒ, hé yī jīn xīlánhuā” (I bought a chicken, two cucumbers, one strawberry, a bunch of bananas, three apples, and a pound of broccoli.)</td>
</tr>
</tbody>
</table>

The six participants whose reports indicated medium levels of depth of processing recognized the classifier and its meaning or association with the noun. In Excerpt 22, the participant was able to associate the measure word ping with bottle, because there is a corollary classifier in English. However, the measure word for book causes more hesitation as the participant passes through the feedback think aloud session.
Table 31. *Examples of Direct WCF with Medium-Level of Depth of Processing on Classifiers*

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>22. “Liǎng píng jiàngyòu” (Two bottles of soy sauce). Oh I see, two bottles of soy sauce. <em>Sān bèn shū</em> (three books), shu is books, ben is the measure word? Umm…ok, next…”</td>
</tr>
</tbody>
</table>

One participant exhibited a high-level of depth of processing on classifiers, taking the time to reflect on the sentence and extracting the measure word. In Example 24, they first recognized the error and then reflected on the difference classifiers used for the different fruits.

Table 32. *Examples of Direct WCF with High-Level of Depth of Processing on Classifiers*

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>23. “Wait, I need to go back. “Hé liàng jīn yú, yu?” (And two pounds of fish), oh, that’s my measure word. “Hé yī hé cāoméi.” <em>And one box of strawberries</em> I didn’t add a measure word there. “Hé yī duì? (And a pair) Is that a measure word for Xiāngjiāo? Hé yī duì Xiāngjiāo. (Bananas? And a pair of bananas)”</td>
</tr>
<tr>
<td>24. “I didn’t know the measure word for box and a bunch. I see. Yī hé cāoméi, Yī bā xiāngjiāo (One box of strawberries, one bunch of bananas)…These things have different measure words…”</td>
</tr>
</tbody>
</table>

For those 7 participants who received indirect WCF with metalinguistic explanations exhibited low depth of processing, their think-aloud responses were similarly repetitions of the text, sometimes repetitions of the error. There is limited acknowledgment of the indirect WCF with metalinguistic explanations. For example, in Excerpt 25, they participant acknowledged the error code, but did not seem to interpret the significance of it.
Table 33. Examples of Low-Level of Depth of Processing on Classifiers: Indirect WCF with Metalinguistic Explanations

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>25. “So many MW, emm… Cài méi, or Cǎoméi?” (Strawberries?)</td>
</tr>
<tr>
<td>26. “So again, the measure word. With liǎng gè sān gè. (Two things, three things) The next error…”</td>
</tr>
</tbody>
</table>

The participants who processed indirect with metalinguistic explanations at the medium level acknowledged the feedback code and its meaning. For example, in Excerpt 27, the participant notes the classifier feedback and repeats the sentence. Similarly, in Excerpt 28, they see the note about word order, but did not fully digest the application to their sentence and the placement of the classifier.

Table 34. Examples of Medium-Level of Depth of Processing on Classifiers: Indirect WCF with Metalinguistic Explanations

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>27. “Oh, there is another MW, just measure word, it should be Sān gè shū.” (Three books)</td>
</tr>
<tr>
<td>28. “This is box, then WO is, oh, word order….thinking about it as yī gè box of, the word order….So píng is bottle, hé is box…yeah…”</td>
</tr>
</tbody>
</table>

Five participants in the indirect WCF group achieved high depth of processing on classifiers. These participants’ think-aloud responses involve a consideration for what the classifier’s category is and try to ascribe a meaning for it in English. There is a meaningful attempt to systematize the classifier in their knowledge. In Excerpt 30, even though they are struggling with the correct term, the participant is considering the different classifiers based on shapes as a result of the multiple indirect WCF codes on the page.
Table 35. Examples of High-Level of Depth of Processing on Classifiers: Indirect WCF with Metalinguistic Explanations

<table>
<thead>
<tr>
<th>Excerpt</th>
</tr>
</thead>
<tbody>
<tr>
<td>29. “Yes, umm… the MW is measure word. umm… it should be liàng gè? It’s the number measure word. It’s counting the number of the objects. You have to have that there. So it would be “liàng gè jiàngyóu.” (Two soy sauces)</td>
</tr>
<tr>
<td>30. “So Jīn for long and round things? Umm? Then you have long and flat things? Oh, it’s gēn. What? Umm…gen for long and round. Cucumber? Wait, tiao? Like pencil. Here are more things. I’m looking for the common themes and an easier way to remember them…Maybe it’s just like the Chinese characters.”</td>
</tr>
</tbody>
</table>

4.2.3 Depth of Processing Over Time

To further analyze how participants processed WCF, two participants’ think-aloud responses, for le and for classifiers, were analyzed for depth of processing at each of the first ten errors in their drafts. The two participants pseudonyms and treatment groups are Jim (indirect group) and Mu (direct group). The plots of their depth of processing, below, show different patterns in how participants process direct and indirect WCF with metalinguistic explanations from Writing Task 1.
Table 36. *Overview of Detailed DoP Analysis Participant Background with Le Examples*

<table>
<thead>
<tr>
<th>WCF</th>
<th>Participant ID</th>
<th>Wring Task 1 Accuracy</th>
<th>Depth of Processing</th>
<th>Revision of Writing Task 1</th>
<th>New Writing Task 2 Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect WCF</td>
<td>Jim (pseudonym)</td>
<td>0%</td>
<td>High</td>
<td>78%</td>
<td>88%</td>
</tr>
<tr>
<td>with metalinguistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>explanations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Think-aloud example:

**High:** Wǒ xué xí, emm, it’s p-le1, it’s umm… particle *le* usage one. Error is wǒ qù School. And the correction is wǒ qù *le* School. So I guess I should put *le* after xué xí. So wǒ xué xí *le* zhōng wén liǎng nián. Participle *le* usage two. Change of state. hǎo de, then hǎo *le* So I should put *le* after nián? Umm…

**Low:** Qùnián, wǒ…it’s particle *le* usage again. Just put *le* before Afghanistan…

<table>
<thead>
<tr>
<th>Direct</th>
<th>Mu (pseudonym)</th>
<th>0%</th>
<th>High</th>
<th>100%</th>
<th>30%</th>
</tr>
</thead>
</table>

Think-aloud example:

**Low:** Zuo, *zuole* shenme? So…Ok, I was try to say what did you do but I don’t know… Ok, next one.

**High:** Ok! Let’s see. I understand most of the sentences, pinyin and Chinese words. I also appreciate the smaller corrections as well. Especially the *le* being added many times. Events in the past. I need *le* there. And some words…

Both participants scored 0% on *le* usage accuracy on Writing Task 1, achieved high depth of processing during the think aloud on WCF. Participant Jim from the indirect WCF with metalinguistic explanations group reached 78% accuracy in the revision, and participant Mu from the direct group reached 100% accuracy in the revision. In the delayed posttest, participant Jim maintained 88% accuracy and participant Mu fell to 30% accuracy.
Data for the depth of processing in *le* for the direct group participant Mu and the indirect WCF with metalinguistic explanations group participant Jim are presented in the figure below.

![Graph showing depth of processing scores for direct and indirect groups](image)

**Figure 20. Le Depth of Processing on the First Ten Errors**

In the think aloud data, the direct participant Mu’s data suggest a very low depth of processing as he quickly moved through the first six errors with barely any acknowledgment of the *le* error, such as “Zuo, zuo le shenme? So…Ok, I was try to say what did you do but I don’t know… Ok, next one.” By the time Mu got to the seventh and eighth error, he began to notice there was a high number of *le* errors, but his comments still do not indicate a real understanding of the nature of the error until the tenth error in the data set. At this point, participant Mu’s data suggest he was processing at a higher depth as he began a deductive process to try to interpret what the ten errors on this pattern mean, as he mentioned at the end of the think-aloud “Ok! Let’s see. I understand most of the sentences, pinyin and Chinese words. I also appreciate the smaller corrections as well. Especially the *le* being added many times. Events in the past. I need *le* there. And some words…” It was as if at this point participant Mu realized that there was enough
information to try to understand the pattern, but would not try to attempt this earlier without
enough information to deduce the rule. All of this was dependent on the participant Mu noticing
the error and potential for a patterned rule, and it seemed that the participant Mu could have
easily continued to flip through the direct WCF without truly noticing the significance of the le
errors.

The indirect feedback participant Jim started at a greater depth of processing and
maintained a higher level in the first few errors. The think aloud showed the participant Jim
referring to the metalinguistic feedback sheet and thinking much more about the pattern. It took
Jim the first four errors to fully process the meaning of the errors, examples, and the
metalinguistic feedback, but once Jim achieved an understanding of the rule at error four, he
subsequently only acknowledged that errors five through ten map to the same le pattern and
processed these le feedback at the low level by simply adding the corrections.

These two plots based on the think alouds are only anecdotal examples, but the plots
suggest that for a pattern-based rule, the two types of WCF may have opposing trajector
ies. Direct WCF may require the participant to piece together meaning from a large volume of
positive evidence, while also placing them at risk of not succeeding in their deduction. The
indirect WCF with metalinguistic explanations provides much more rule information via the
feedback sheet and the metalinguistic explanation may guide the student to process the rule on
fewer errors.

Similar data for classifiers is presented in the Figure 21 below. Both the direct and
indirect participants processed the feedback in a similar pattern and their think-aloud responses
reflected a similar reaction to the different feedback. Both participants had the basic
understanding of the classifier system and scored 25%, and 30% accuracy respectively in
Writing Task 1. Both participants greatly increased in classifier accuracy on the revision (80% and 100%). However, neither of them maintained the accuracy improvements and had lower accuracy scores in New Writing Task 2, perhaps due to the perceived classifiers’ level of difficulty.

Table 37. Overview of Detailed DoP Analysis Participant Background with Classifier Example

<table>
<thead>
<tr>
<th>WCF</th>
<th>ID</th>
<th>Wring task 1 Accuracy</th>
<th>Depth of Processing</th>
<th>Revision of Writing Task 1</th>
<th>New Writing Task 2 Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indirect with metalinguistic explanations</td>
<td>Jim</td>
<td>25%</td>
<td>High</td>
<td>80%</td>
<td>18%</td>
</tr>
</tbody>
</table>

Think-aloud example:

High: So jǐn is not for cucumber? I guess gēn here is for the long and round thing. So, carrots would be used gen as well?

Low: Yī jǐn huáng guā. Emm… all of the measure words are very difficult, haha kidding, thanks for the feedback.

| Direct | Mu   | 30%   | High | 100% | 22% |

Think-aloud example:

High: Liǎng píng jiàngyóu, I see ping for soy sauce, hé sān běn shū,. shū is book, so běn is the measure word for books.

Low: Yeah, yī zhī jǐ, cǎoméi, xī lánhuā, jǐ means Chicken. Oh, cǎoméi, not tǎoméi
In the first five errors, both participants Jim and Mu reacted to the classifier and notice the correction, while trying to recall the object category reflected in the classifier.

Participant Mu from the direct group commented “Liǎng píng jiàngyóu, I see ping for soy sauce, hé sān bèn shū, shū is book, so bèn is the measure word for books…”

Participant Jim from the indirect-with-metalinguistic-explanations group expressed similar understanding of classifiers in his think aloud session “So jīn is not for cucumber? I guess gēn here is for the long and round thing. So, carrots would use gen as well …”

As participants Jim and Mu progressed through the errors, however, they eventually reached a point of saturation where the classifiers seemed too irregular or too overwhelming in the course of reviewing the assignment. There was a drop-off after error five because of a particularly surprising or challenging classifier (classifier zhi for the chicken) that seemed to

Figure 21. Classifier Depth of Processing on the First Ten Errors
have pushed both participants to give up on trying to process each classifier as a categorical descriptor, and subsequently, the participants just skimmed through the classifier errors.

Participant Mu chose to pay attention to the wrong words after the classifier such as chicken and the misspelling of strawberry. Participant Jim jokingly indicated that there were too many classifiers and there was not a clear rule to use them correctly by saying “yi1 jin1 xi1lan2hua1. Emm… all of the measure words are very difficult, haha kidding, thanks for the feedback…”

Although these are only two anecdotal examples, this similarity in pattern suggests that the feedback in this area could lose its potency when it seems overwhelming. The breadth of classifiers in Chinese may bring out the frustrating aspects of foreign language learning—the rote memorization and the linguistically foreign concepts—and lead to an emotional response that reduces depth of processing.

4.2.4 Summary of Research Question 2 Results

Overall, the data in this subsection showed that indirect WCF with metalinguistic explanations led to predominantly high DoP for the particle le and mixed DoP for classifiers. The excerpts from participants’ responses provide insight into how some participants in the indirect-WCF-with-metalinguistic-explanations group engaged in an internal problem-solving discussion that may help explain some of the accuracy outperformance identified in the previous section. For direct WCF, participants displayed similarly low DoP on both particle le and classifiers. The simple repetitions of the low DoP excerpts make it clearer why concepts are unlikely to sink in and performance improvements do not carry into the New Writing Task 2 results.
4.3 Research Question 3

The third research question asked “How do low-proficiency L2 Chinese writers engage with the two different types of WCF (direct vs. indirect with metalinguistic explanations)?” Ellis (2010) provided a conceptual framework for student engagement with teacher’s written corrective feedback which included behavioral, cognitive, and affective engagement. This framework was employed to explain participants’ engagements with direct WCF and indirect WCF with metalinguistic explanations. The analysis of the 38 participants’ written accuracy on particle le and classifiers in Writing Task 1, the Revision of Task 1, and New Writing Task 2 was used as an indication of participants’ behavioral engagement. Analysis of the 38 participants’ depth of processing in the think-aloud session provided an indication of participants’ cognitive engagement. At the conclusion of the course, the 38 participants filled out questionnaires and attended online semi-structured interviews comparing direct written corrective feedback and indirect written corrective feedback with metalinguistic explanations. The survey data was used as an indication of the participants’ affective engagement. The previous sections on Research Question 1 and Research Question 2 already presented accuracy analysis and depth of processing. The results summarized here were gathered from a questionnaire and post-data collection semi-structured interviews, as well as the teacher focus group.

4.3.1 Student Engagement

As reported above, both the direct and indirect-with-metalinguistic-explanations groups were successful in improving their performance on the Revision of Writing Task 1; however, only participants in the indirect-with-metalinguistic-explanations group maintained the accuracy in usage of le after two weeks. Based on the qualitative data gathered on the student interviews and
questionnaire, it appears that the indirect-with-metalinguistic-explanations group had a higher behavioral engagement than the direct feedback group.

In the student interview and questionnaire about their perceptions of WCF, there were two items related to student behavioral engagement that asked participants to consider labeled examples of direct and indirect feedback. The first set of questions was designed to address which type of WCF was easier to revise and apply in future writings. The second set of questions asked participants to choose the type of WCF that was most beneficial for their future learning, and was designed to explore which type of feedback would make participants notice the underlying rules tied to metalinguistic awareness of their grammatical development. On the WCF questionnaire, after seeing examples of direct and indirect WCF, the majority of participants (85%) expressed that direct WCF was easier for them to correct the errors; however, the majority of participants (92%) indicated that they preferred indirect feedback with metalinguistic explanations.

Interestingly, in the interviews, indirect WCF participants expressed confusion and hesitation when they tried to correct the errors based on metalinguistic codes and examples. They were not confident about their error corrections and hoped to check their corrections against the teacher’s feedback. Some of the indirect WCF group participants also mentioned it was their first time seeing metalinguistic codes and examples. It took more time and effort for them to figure out the answers with indirect WCF than with more traditional direct feedback methods. When asked to further comment on their preferences for feedback to support their future language learning and writing development, while these participants expressed their preference for quick and easy WCF, they commented that indirect WCF was more beneficial.
Participant Emily (pseudonym) from the indirect WCF group seemed somewhat indifferent between direct and indirect feedback, but suggested that indirect allowed her to focus on the areas she was struggling with during the one-on-one sessions:

I think both types of feedback would be very helpful! I like the feedback I received more [indirect]. I never paid attention to *le* and I didn’t know *le* could be used with verbs to indicate past tense. Studying with the codes and examples was very beneficial and helped me remember to use *le* in the future. I also like that the indirect feedback gave an opportunity to figure out the correction. I was less embarrassed to discuss my draft with you. I was working hard on the first draft and didn’t realize there were so many errors. I could fix them on my own on most of the time, but it was probably a good use of your time to work on the difficult ones with me…

Participant Fan (pseudonym) highlighted how indirect WCF with explanations helped her learn from her mistakes, while suggesting direct feedback can be overwhelming or shallow:

[After being shown examples of direct and indirect feedback] I don’t like the direct feedback that I got from my previous language classes. I don’t need a perfect essay or report. I think the purpose [of feedback] is to review the things that I’ve studied. It makes more sense to view my errors with examples, so I can work on [those areas], on the errors with your help. My husband studied this in school and he always just told me the corrections, but this doesn’t help me much. He told me that I have to just memorize them first. Haha…. I disagree. On the contrary, as a beginner, I think I need more discussion and to
work on the errors…umm, it’s important to understand why it was wrong then correct it. Then I can feel like I’m making progress…

Participant Tim (pseudonym) from the direct WCF group said:

I know when I’m making a mistake about le. But I don’t know why and how to use le. I don’t know. I asked the teacher and she didn’t tell me. I hope to study the rule and see more examples. That would be more helpful when I realized there were so many le errors in my mini report.

Participant Yona (pseudonym) from the direct WCF group mentioned similar thoughts and expressed that she would rather work on the errors on her own first. She then compared her corrections with the teacher’s. This participant was multilingual and had experience as an ESL teacher, and was meta-aware of the benefits she was experiencing with the feedback and grammar.

Most of my errors about prepositions and classifiers, umm, I just forgot the Chinese way and put my sentences in the English order. If the teacher would point them out, I think I can fix them on my own, and that would leave a deep impression. I really like how the teacher asked me to say certain expressions again in class, then I knew something was wrong there. Then I realized the location preposition should be placed in front the action and sometimes I just forgot to add classifiers. I feel like indirect feedback with examples is more beneficial and the teacher will work with me patiently.
Although the majority of participants indicated indirect WCF with metalinguistic explanations was more helpful, three participants preferred more direct WCF. Participant Zoe (pseudonym) stated their preference for direct feedback in the beginner setting:

I like being directly corrected, I know I’m a beginner and I should remember a lot of expressions. Some structures are very advanced. The first step is to memorize them. Just like learning radicals is the first step of Chinese character learning. I really like the common Chinese expressions you gave us. So I can use “man man shuo” [speak slowly] or “ting bu dong” [I don’t understand by listening] right away. When I don’t remember, you can just put them there. It’s impossible for me to construct these expressions.

Regarding the participants’ cognitive engagement, the depth of processing ratings from the feedback think aloud recordings revealed a difference between the direct and the indirect-with-metalinguistic-explanations group. The participants who received direct written corrective feedback with metalinguistic explanations reported low DoP, with only 11% achieving high depth of processing on le and 5% reaching high depth of processing on classifiers, while the participants who received indirect feedback with metalinguistic explanations reported high levels of depth of processing, with 78% high on le and 26% on classifiers. The depth of processing data indicated that the direct WCF group did not engage participants in meaningful evaluation of le and classifiers. In comparison, indirect feedback with metalinguistic explanations encouraged participants to engage in higher processing of the abstract metalinguistic codes and underlying grammatical rules. Indirect WCF with metalinguistic explanations had a higher cognitive engagement than the direct WCF group in this study.
Lastly, from the affective perspective, most participants from both direct and indirect-with-metalinguistic-explanations groups were quite positive about WCF. On their feelings about teachers’ WCF on the questionnaire, participants reported the following impressions:

“Mixed feelings” (Indirect WCF with metalinguistic explanations)

“Happy? be real -- satisfied that although I was wrong, maybe I was close” (Direct WCF)

“I was happy to receive the feedback and appreciated not having to guess any more on my own.” (Direct WCF)

“The more the better. I have no problem with corrective feedback.” (Indirect WCF with metalinguistic explanations)

“Very happy” (Direct WCF)

“Happy and nervous (if I make too many mistakes)” (Direct WCF)

“Happy to get the feedback” (Indirect WCF with metalinguistic explanations)

“Happy and understanding” (Direct WCF)

In the interviews, most of the participants expressed appreciation for WCF:

“Thank you! I like feedback! The difficult part is not the corrections. I think Chinese is hard to learn either way!” (Indirect WCF with metalinguistic explanations)

“Very happy to get corrections! Corrective feedback that identifies where the mistake is, but forces me to work out why it is a mistake and how to correct it
seems the most effective. I like your feedback!” (Indirect WCF with metalinguistic explanations)

“Yes, I like the corrections. I have actually learned best through mistakes. I tend to remember the language better after I have really messed it up.” (Direct WCF)

“I like it, it gives me something to model on--it is clearly and unambiguously correct.” (Direct WCF)

“I really like the new method. Very nice! I like the very specific feedback and the ability to learn from the explanations. Great idea!” (Indirect WCF with metalinguistic explanations)

At the same time, some participants from both feedback groups exhibited frustrations with the type of feedback they received. Participants in the direct feedback group indicated that the direct feedback they received might have induced anxiety due to the overwhelming amount of errors being highlighted in their drafts and that this potentially interrupted the learners’ processing of the feedback:

“I should study hard! I’m so sorry that I made so many errors…” (Direct WCF)

“I know there are many classifiers, but I can’t correctly use all of them and there is no rule, it’s just impossible to use the classifier system. Maybe it’s ok to just use ge, the general one if the teacher is fine with me.” (Direct WCF)
“I’m so sorry. Mary was very patient. Maybe I tried to express myself in English and it didn’t work out well. Only the subject ‘wo’ is correct. There are so many new things I don’t understand. Maybe I studied them before and I’ll pay attention to them more, such as the tense? I still don’t know how to describe things in the past. I’m also struggling with the speech order in Chinese…” (Direct WCF)

Participants from the indirect-with-metalinguistic-explanations group said that it took a long time for them to get accustomed to and then work on the metalinguistic codes and examples. Participants from the indirect WCF group commented:

“This method (indirect with metalinguistic explanations) is very confusing. A little less my style, but understandable -- as grammar content deepens, it’s possibly hard to keep doing it.” (Indirect WCF with metalinguistic explanations)

“The codes and examples are a lot of work to get corrections. I’m bad at and nervous about tones. That’s why I didn’t put any tones there. I don’t know how to write them. I hope I can get more direct corrections and just read the draft with me in class. I also need to go back to my textbook to check my corrections. I don’t know how to correct myself and I need to study hard! I hope the teacher would just tell the answers and we can practice them in a conversation. That would work better for me. I should study hard and have a lot of repetitions.” (Indirect WCF with metalinguistic explanations)
4.3.2 Summary of Research Question Three Results

In conclusion, regarding affective engagement, both direct and indirect WCF with metalinguistic explanation feedback groups demonstrated positive engagement with WCF. Some participants in the direct group reflected anxiety when their compositions were full of direct corrections. Participants in the indirect-with-metalinguistic-explanations group exhibited frustration and confusion when they attempted to correct errors based on the metalinguistic explanations. With regard to cognitive engagement, the indirect-with-metalinguistic-explanations group had higher engagement because most participants processed the target structures at high DoP. Regarding behavior engagement, the direct group had a higher engagement in the Revision of Writing Task 1, and the indirect-with-metalinguistic-explanations group had a higher engagement in the delayed posttest New Writing Task 2. Most participants preferred more detailed feedback, especially so they could better understand where they went wrong and how to improve. A minority of participants liked when teachers directly told them the correct answer, particularly when tied to tone and pronunciation, or when the required grammar exceeds the participant’s language ability.

4.4 Research Question 4

To answer Research Question 4 “What are teachers’ perspectives about the two different types of WCF (direct vs. indirect with metalinguistic explanations)?” The researcher engaged in a video-chat focus group with the other three teachers in the course: the teacher that participated as an instructor in the study and two teachers that taught other sections of the program and observed portions of the study.
When asked which type of feedback (direct or indirect with metalinguistic explanations) was easier to deliver to students, the teachers disagreed. Mary (pseudonym) stated that she preferred providing direct feedback due to the ease of delivery and the positive evidence that it allows her students to move on to the revision. She stated: “You just put the correct answers there and move on.” She indicated that explicit feedback became frustrating: “When I try to give indirect metalinguistic feedback with codes and examples—even I don’t remember all the codes. I have to check back and forth. Sometimes, one sentence has many errors and it’s difficult to put all the codes there. Why not just give the correct answers?” Mary also stated some students who received indirect feedback with metalinguistic explanations would say: “I know it’s not correct, and I try to correct myself. How would I check my assumptions?” This frustration affected her perceptions on what type of feedback was most appropriate for students.

The author-researcher explained, “Yes, sometimes students made multiple errors in one sentence. We always have to choose which way to correct them. I think the indirect WCF with metalinguistic explanations places the focus on the correction that is on the feedback sheet. Maybe since I designed the feedback sheet, I found it easier to give codes.” Another teacher, Lucy (pseudonym), who did not participate in the study, supported the author-researcher’s perspective, “I think many students try to say complicated things when they haven’t reached the appropriate proficiency level. If we can give direct corrections, then it kind of encourages this kind behavior. I think this method [indirect with metalinguistic explanations] will ask them to refocus on what the students have learned and not things that students haven’t studied. It’s easier for us as teachers to teach in the end…”

On the teachers’ perspectives of their own students, the author-researcher noted she found it interesting that so many students wanted metalinguistic codes and examples for their
errors. She said: “A lot of students asked me ‘how do you say this or that in Chinese’ and they didn’t give me time to explain how to build these expressions in class. I didn’t expect for them to like to be corrected with metalinguistic codes and examples.” Another teacher who didn’t give feedback treatments indicated a similar view from her teaching experience. She mentioned a similar preference and highlighted the mismatch between students’ preferences and the real work load constraints of a teacher with a limited amount of time to grade assignments: “I think it’s natural that students prefer more detailed feedback, however, our time is limited and we only have 10 minutes to give feedback on students’ assignment each week. We want to give more detailed metalinguistic feedback, but we couldn’t.” Other teachers agreed with this comment and Mary added “when I compare my direct WCF and indirect WCF with metalinguistic explanations, I’m worried that my students think I’m lazy and didn’t spend time to correct their drafts. The direct WCF group really likes my feedback and they feel I work hard too.” The author-researcher shared her disagreement: “We have time constraints, and many of my students in the indirect-WCF-with-metalinguistic-explanations group really appreciate the metalinguistic codes and examples. They felt they shouldn’t waste time in class so they felt bad if it took them very long to figure out certain expressions. They hope to have as much oral practice as they can. However, they can look at the metalinguistic codes and examples at home and take as long as they want to figure out the correction. They commented they make these expressions as their own and are able to produce them in the future. Otherwise, they won’t remember them.”

When the teachers were asked to interpret the quantitative outcomes of the study on le and classifiers, the three other teachers expressed surprise and interest. Mary stated: “I thought both groups would have similar outcomes, not improve much at all, since maybe the direct group already has the answers, especially with classifiers, they should perform better!” The teacher-
researcher explained her rationale for why there was an observed difference between the two
groups at the end of the study: “I got a feeling the indirect-WCF-with-metalinguistic-
explanations group not only paid attention to the corrections, but also processed these corrections
more deeply than the direct group. But we’ll confirm my guess with depth of processing data
later (explaining depth of processing to other teachers).” Mary indicated that her perception was
also connected with the proficiency level of the learners. Given that the participants were low-
beginners, she stated she felt the more metalinguistic feedback increased student anxiety. She
said, “I think my students are pure-beginners, they haven’t mastered the Chinese grammar
systems yet, so many special grammar structures. They even had a hard time memorizing basic
vocabulary and expressions, such as Nǐ hǎo (hi), guìxìng (family name is) versus gāoxìng
(happy). I don’t want to add extra burden to their learning. They don’t know what aspect or
classifiers in English is. I guess…I didn’t know these words in Chinese and English before…I
thought just giving them the direct corrections might be better for these beginners.”

4.4.1 Summary of Research Question Four

Results from the teachers’ focus group interview indicated a mismatch between the
teachers’ initial beliefs about WCF for grammar learning and the outcomes of the study as shared
with them during the session. They had all seen the material and treatments, and so had clear
preconceptions about what the results would look like. In the post-data collection interviews, the
teachers were surprised to find that the most effective form of feedback for both le and classifiers
were indirect with metalinguistic explanations. Excluding the author-researcher, the other
teachers felt nervous about implementing indirect WCF with metalinguistic explanations.
CHAPTER V: DISCUSSION

The primary goal of the current study was to examine the effects of unfocused direct WCF versus unfocused indirect WCF with metalinguistic explanations on L2 learning of Mandarin le and classifiers in a 14-week synchronous computer-mediated communication task-based curriculum. The secondary goal was to explore depth of processing by type of WCF (direct and indirect with metalinguistic explanations) via think-aloud protocols. The third goal was to investigate students’ WCF engagement. The fourth goal was to understand teachers’ perceptions and preferences concerning WCF and the learning of le and classifier. This chapter presents discussions in relation to previous research as summarized in Chapter Two.

5.1 Research Question 1: What is the Effect of Different Types of WCF (Direct vs. Indirect with Metalinguistic Explanations) on Low-proficiency L2 Chinese Writers’ Accuracy with Le and Classifiers?

Accuracy results on both particle le and classifiers demonstrated that the direct WCF group statistically outperformed the indirect-WCF-with-metalinguistic-explanation group on the Revision of Writing Task 1, but this difference in performance was not maintained on New Writing Task 2. While both groups’ accuracy decreased significantly from the Revision of Writing Task 1 to New Writing Task 2, only those in the indirect-with-metalinguistic-explanations feedback condition maintained a significantly improved accuracy in usage of particle le after two weeks in Writing Task 2 when compared to Writing Task 1 with a medium effect size.

The current study’s findings support the research that found indirect WCF with metalinguistic explanation more effective than direct WCF (e.g., Evans et al., 2011; Hartshorn et al., 2010; Lalande, 1982; Sheen, 2007; Shintani & Ellis, 2015). One possible explanation from
Lalande’s (1982) analysis on the advantage of indirect feedback suggests that indirect WCF (with or without metalinguistic explanations) triggers error awareness in learners that stimulates the development of writing accuracy through problem-solving techniques. When students get engaged with indirect WCF, they need to spend more time and mental effort to understand the linguistic codes, explanations, and consider applying the information to correct the marked errors. Therefore, indirect WCF may have a more significant beneficial effect on the development of writing accuracy because it requires the student to exert more effort and does not allow them to quickly process or skip the feedback.

The results found in the current study contrast with previous studies that found no difference among different types of WCF (Robb et al., 1986; Semke, 1984; Mubarak, 2013; Caras, 2019). One possible explanation for the different results may lie in the differences in study design. Robb et al. (1986), Semke (1984) and Mubarak (2013) adopted holistic error measurements such as all errors types, or errors per 100 words. Tracking improvement on each type of error is difficult with holistic error measurement, compared to this study’s focus on specific target items. It is also possible that unfocused WCF is too overwhelming for the participants to focus on the target error correction and then produce the corrected form during future writing (i.e., Ellis et al., 2008). This study’s differences with Caras (2019) may arise from the participants’ high prior knowledge of imperfect and preterite in Caras’s study, leaving little room for improvement overall.

Based on the satisfactory performance on le usage two weeks after the indirect WCF with metalinguistic explanations treatment, the findings of the current study provide some support to refute studies suggesting that unfocused WCF does not have a lasting effect (e.g., Bitchener & Knoch, 2008a; Ferris et al., 2013; Kepner, 1991; Mubarak, 2013; Polio, Fleck, & Leder, 1998;
Truscott & Hsu, 2008; Caras, 2019). Other studies have also found longer-term benefits from indirect WCF. Sheen (2007) targeted the acquisition of English articles and found an advantage for metalinguistic explanations, but this superior performance only became evident 2 months later in her delayed post-test. Moreover, language analytic ability was more strongly related to acquisition in the direct WCF with metalinguistic explanations group than in the direct only group. On the other hand, Bitchener and Knoch (2008a, 2008b) found that the difference between direct and indirect WCF with metalinguistic explanations were not statistically significant after a two-month period. However, Bitchener and Knoch’s direct WCF treatment was combined with an oral metalinguistic explanation in the form of five-minute one-on-one conferences. Interestingly, they found that the performance of the direct WCF with metalinguistic explanations group in the delayed post-test declined after seven weeks because of the target item’s overuse. It is a common phenomenon in Instructed Second Language Acquisition research that participants who receive metalinguistic feedback or instruction may overemphasize a pattern—the results in the current study are only applicable to a period of two weeks and may similarly decrease if continued exclusively in this way.

Lastly, the current study contrasts with those studies that found direct WCF more effective than indirect (e.g., Chandler, 2003; Suh, 2010). Chandler (2003) found that the direct WCF treatment outperformed unfocused indirect WCF with metalinguistic notes, however, this result arose in a decidedly unique design. The 36 participants received both direct and indirect WCF treatments, but in different orders. Chandler may have found more evidence for the order of WCF and how different types interact in a mixed scenario.

In the end, all comparisons among these studies’ results should be considered with caution because of major differences in their research designs. They vary in areas such as
proficiency level, WCF treatment type, and accuracy measures. For example, Hartshorn et al. (2010) also added a one-on-one discussion to the indirect WCF treatment. It would be very difficult to parse whether the accuracy improvement in those cases was due to the type of WCF itself or the added discussion.

The quantitative results reported above on the particle *le* and classifiers appear to provide some evidence that supports Ferris’s (1999) suggestion that indirect WCF would be most effective where the errors are “patterned and rule-governed,” and that no single form of correction can be effective for all kinds of errors. The current study chose to compare direct and indirect WCF with metalinguistic explanations on two distinct targets that present different challenges to learners. Particle *le* is a more rule-based concept. It has two salient grammatical features: to address action completion when *le* is added after verbs and to indicate change of the current state when placed at the end of the sentences. Classifiers, on the other hand, do not have clear rules except for some minor categorical associations that can be overlapping or irregular. L1 English speakers and many participants in the current study expressed frustration over the number and arbitrariness of Mandarin classifiers, as noted by, for example, Crosthwaite (2014) and Crosthwaite, Yeung, Bai, Lu, and Bae (2018). Even Chinese language teachers find it is difficult to master many classifiers at once, and it is advised to learn the Chinese classifier system through gradual memorization (e.g., Zong-Jiang, 2004; Li & Thompson, 1981).

In this study, *le*, as the more rule-governed target, benefits from indirect WCF with metalinguistic explanations more than direct WCF. Post-hoc pairwise comparisons analyses revealed that the indirect-with-metalinguistic-explanations group outperformed the direct group on New Writing Task 2 with a mean accuracy difference of 66% and with a large effect size. The indirect feedback with metalinguistic explanations appeared to have created a durable
improvement in performance that remained two weeks after the treatment, whereas the direct
group regressed to a similar level of performance when compared to their initial performance on
Writing Task 1. This significant performance by indirect WCF with metalinguistic explanations
also carried over to the results for classifiers; however, the effect size was small.

The varying results for le and classifiers are expected because ample evidence in SLA
research has demonstrated that linguistic structures vary in their learnability because of
complexity and saliency (Li, 2014). The complexity of linguistic structure may also play an
important role in WCF effectiveness. The superior results for le as compared to classifiers are
also in line with previous research that shows rule-governed grammatical structures are easier to
correct than more idiosyncratic grammatical structures (e.g., Bitchener, Young, & Cameron,
2005; Bitchener et al., 2010; Shintani et al., 2014). For example, Bitchener et al. (2005) reported
accuracy gains for the rule-governed past tense and definite articles, but not for rote prepositions.
Shintani et al. (2014) found only the group that received indirect WCF with metalinguistic
examples maintained accurate production of the hypothetical conditional tense on a new delayed
post-test. Future research might want to investigate different linguistic structures as variables and
run statistical analyses to further untangle the relationship among different WCF types and
different linguistic structures.

5.2 Research Question 2: How do Low-proficiency L2 Chinese Writers Process the Two
Different Types of WCF (Direct vs. Indirect with Metalinguistic Explanations) on their Written
Production?

The depth of processing ratings from this study show a clear difference between indirect
and direct WCF that is supported by the accuracy results and student interviews. Overall, on both
le and classifiers, the direct WCF group was weighted toward lower levels of DoP, with 78% on
le and 63% on classifiers. This is sharply contrasted with the indirect group, where the majority (74%) exhibited higher levels of DoP and only 5% displayed low DoP on le. The qualitative results from think-aloud responses and student interviews reinforce that direct WCF did not engage participants in meaningful evaluation of le and classifiers. Since students were allowed to reference the WCF when doing their Revision on Writing Task 1, direct WCF participants achieved almost perfect accuracy scores on the revision. However, given that this type of feedback essentially spoon-fed them the answers, they did not appear to have engaged in much deeper processing. Most direct WCF participants focused their reflection on lexical errors or tones because they were more salient than the grammatical points—an entire lexical correction can be captured in the WCF, but the grammatical point must be inferred. This data can provide a plausible explanation to account for the sharp decrease in performance by the direct WCF participants’ performance on New Writing Task 2 because they had not internalized the pattern or rules inherent in the feedback. On the other hand, the indirect WCF group was able to maintain accuracy in le because the indirect WCF essentially required them to engage in deeper processing of the abstract codes and metalinguistic feedback. Indirect WCF participants focused on grammar rules in their think-aloud responses and engaged in problem-solving activities while interpreting the feedback. This deeper processing appeared to result in greater retention after two weeks.

Other studies investigating WCF and depth of processing employed Schmidt’s (2001) “levels of awareness.” Schmidt describes the levels as: 1) at the level of noticing the feedback, and 2) at the level of understanding the feedback, where learners intuit the underlying rule of the target. Studies have argued that higher levels of awareness can lead to better performance, especially when reaching the level of understanding (e.g., Leow, 2001a, 2019; Hsieh, Moreno, &
Leow, 2015; Suzuki, 2017). In particular, incorporating Schmidt’s levels of awareness and other theoretical underpinnings in SLA, Leow (2019) proposed the feedback processing framework to account for cognitive processes employed during the stages of addressing WCF: (1) learners receive L2 information via WCF based upon their own L2 written product; (2) learners cognitively process WCF in relation to their prior knowledge; (3) the further processed information is registered or restructured in the knowledge system; and (4) when there is a new opportunity to produce the output, learners may produce old output or new or modified output dependent upon how the feedback was or was not cognitively processed. Besides potential levels of awareness, this framework addresses depth of processing, usage of prior knowledge, hypothesis testing, and rule formulation in WCF effectiveness.

The findings of the current study provide supporting evidence for Leow’s feedback processing framework and the argument that higher levels of awareness can lead to better performance on subsequent tasks. The direct WCF often guided participant’s attention to the target at the level of noticing. For example, one participant noted that “The next sentence I try to say is ‘I studied for two years.’ So Wǒ xué xí le liǎng nián le (I have studied Chinese for two years). Umm… I don’t know this. I need some help on this. Maybe we’ll go back to this later.” This participant saw le was directly added to his sentence, but he did not choose to process it further and did not interact with his own established internal grammar system. According to Leow (2019), this participant in the direct WCF group only processed at a low level and only arrived at the level of noticing on his le feedback. Information obtained at this surface level noticing may be easily lost if the participant does not make a link between the new information from the direct WCF and his own prior knowledge. By not further processing the WCF provided, the opportunity to employ processes such as hypothesis testing and potential rule formulation is
reduced for this feedback to be restructured in his internal system. By only copying and pasting the corrections in the revision, participants from the direct WCF group appeared not to have processed the WCF deeply enough to retain their *le* accuracy on the delayed post-test (New Writing Task 2).

On the other hand, in line also with the postulations of Leow’s feedback processing framework, indirect WCF with metalinguistic explanations appeared to have pushed participants to process feedback at a deeper level leading to a higher level of awareness in many instances (level of understanding). This is also consistent with Storch and Wigglesworth (2010)’s argument. For example, in the current study, one participant stated in the think aloud session “The first error where you have *le* is in the sentence ‘Rènshi…’ I guess *Nana hěn gāoxìng* (very happy). That type of error on the feedback sheet says *le* usage expressing completion of action. You gave examples and how to correct it … so event in the past, or the English past tense. I wrote, ‘It was nice to know Nana.’ So it would be *Rènshíle hěn gāoxìng* (It was nice to meet Nana). Is that correct? Next one…” Participant 15 from the indirect WCF group saw the code “P-le1” and checked the explanation and examples of *le*. First, this participant noticed there was an error about *le* in the sentence, and referenced the code sheet. The indirect WCF with metalinguistic codes and examples pushed him to further process the *le* grammar beyond the level of noticing. After studying the metalinguistic codes and example, this participant incorporated new information (*le* means action completion in Chinese) and interacted with his own prior knowledge (English past tense). Then he added the new information about *le* into his internal grammatical system pertaining to the use *le* to describe past action in Chinese. The high depth of processing together with a high level of awareness at the level of understanding could provide a plausible explanation for his retention of the *le* accuracy in the delayed post-test.
The depth of processing results demonstrated that direct WCF can draw learners’ attention to the target structure mostly at the level of noticing, whereas indirect written corrective feedback encourages learners to process the feedback at a deeper level, frequently reaching the level of understanding, by pushing learners to figure out the correction through a problem-solving activity. When learners move beyond the level of noticing and arrive at level of understanding, this kind of depth of processing on rule awareness facilitates better performance, especially on long term gains.

Several other studies have also reported low depth of processing associated direct WCF while indirect WCF promoted deeper levels of processing (e.g., Caras 2019; Ferris, 2001; Sheen, 2007; Wigglesworth & Storch, 2012). Caras (2019) found that 64% of participants in her study receiving direct WCF processed ser and estar at low levels, while the participants receiving indirect WCF with metalinguistic information processed at the medium level. However, unlike the current study, both Caras’s direct and indirect groups did not maintain their accuracy gains after two weeks.

This study’s results also appear to contradict the findings of Cerezo et al. (2019). They found their direct WCF group reported a high depth of processing, and engaged in self-analysis of their errors, providing explanations on the errors. Their indirect WCF participants had low depth of processing, mentioning possible error corrections or error category. Some of them processed at a deeper level, but they left their errors uncorrected.

There are two possible explanations for this contrasting result. First, the participants in Cerezo et al.’s (2019) study were not pure beginners; their proficiency level was International English Language Testing System (IELTS) Band 4 indicating that they were intermediate level learners on the ACTFL scale. In other words, these participants already had basic linguistic
competence for daily situations, and likely had the foundational concepts for many of the grammar patterns that they encountered. They might have been able to apply this foundational knowledge to place direct WCF within the patterned framework they had already developed.

A second factor lies in the differences in research design. To assess depth of processing, Cerezo et al. (2019) employed a separate written languaging session where participants had to write down their reflections on the corrected task. Since the direct WCF group had the correct usage provided to them via the direct corrections, it was easier for them to deduce the explanation and the rules in the session. Their indirect WCF group was only provided with metalinguistic codes and their meanings—without any examples of correct usage—making it quite difficult for the indirect WCF group to interpret their error and spontaneously recall the correct pattern. Influenced by this depth of processing measurement method, the indirect WCF group in Cerezo et al. (2019) likely had a bimodal distribution for low and high depth of processing.
5.3 Research Question 3: How do Low-proficiency L2 Chinese Writers Engage with the Two Different Types of WCF (Direct vs. Indirect with Metalinguistic Explanations)?

Student engagement and teachers’ perspectives on WCF can generate contextualized knowledge to support the effective use of WCF in the language classroom. In the focus groups and interviews, using WCF to correct a given error was a common topic of discussion in participant feedback. The majority (85%) of student participants found direct WCF easier for them to digest and apply immediately to revise their tasks, consistent with the depth of processing analysis, above. Surprisingly, the indirect WCF with metalinguistic explanations participants reported that the metalinguistic codes and additional examples caused confusion and hesitation. It was not immediately apparent to them how to apply this feedback and deciphering the codes and examples created stress. This finding matched previous research that students found indirect WCF more difficult to work with (e.g., Chandler, 2003).

Many studies (e.g., Lee, 2008; McMartin-Miller, 2014) have reported that students naturally expect teacher’s WCF and are happy to receive it, even if the WCF does not result in positive learning experiences. In the current study, the majority (92%, n = 38) of participants preferred explicit feedback upon making an error in a writing task. They hope to understand not only what was wrong, but also why and how to get it correct on their own. Therefore, explicit feedback with metalinguistic codes and examples was popular. This finding echoes previous research (e.g., Ferris & Roberts, 2001; Ferris et al., 2000; Hedgcock & Lefkowitz, 1994; Komura, 1999; Leki, 1991; Radecki & Swales, 1988; Rennie, 2000), demonstrating L2 students’ views toward teacher feedback on their written errors. The most popular WCF method for teachers was underlining with description and examples on how to correct the error, then followed by direct corrections.
Most participants (92%) expressed that indirect feedback with metalinguistic explanations was more helpful than direct WCF, showing that the participants seemed accurately self-aware of what was driving better learning outcomes. Indirect feedback with metalinguistic codes and examples would push students to think more deeply and figure out the corrections on their own. This result supports the accuracy and depth of processing results showing the indirect WCF group improved more than the direct WCF group on the delayed second writing task. Depth of processing data from both groups demonstrated that indirect WCF gave students an opportunity to process le at a high level of understanding.

Only three participants considered direct feedback better, noting that they felt their proficiency was too low to understand complicated grammatical structures. Most participants in the direct group indicated that direct feedback was not very beneficial because the direct WCF they received might have induced anxiety due to the overwhelming amount of errors being highlighted in their drafts. This feedback burden might have potentially interrupted the learners’ processing of the feedback. Early research reported WCF could elicit negative emotions for similar reasons (e.g., Semke, 1984; Truscott, 1996). Students might find too much feedback as discouragement, as demonstrated in prior work comparing types of WCF (e.g., Ellis et al., 2008; Sheen et al., 2009). For example, unfocused WCF caused frustration and confusion among students in some studies, damaging their motivation and confidence about writing (Ferris, 2003, 2011; Truscott, 2001). Other studies demonstrated a wide range for WCF’s effects on students’ mental state (e.g., Mahfoodh, 2016; Zhang & Hyland, 2018), describing feeling of accomplishment (Ferris, Liu, Sinha, & Senna, 2013), self-confidence (Storch & Wigglesworth, 2010), disappointment (Hyland, 1998), frustration (Zheng & Yu, 2018), indifference, relief, or even excitement (Han & Hyland, 2015). These dynamic emotional reactions toward WCF could
influence students’ understanding of WCF, their revision behaviors, and other related L2 writing and L2 learning strategies (Han & Hyland, 2015; Ellis, 2010a; Zheng & Yu, 2018). Future empirical studies may want to focus on students’ emotional reactions to different types of WCF throughout the revision process, considering how these emotions may add cognitive load or affect processing capacity.

5.4 Research Question 4: What are Teachers’ Perspectives about the Two Different Types of WCF (Direct vs. Indirect with Metalinguistic Explanations)?

Results from the teacher focus group indicated a mismatch between the teachers’ initial beliefs about the effectiveness of WCF for grammar learning and the outcomes of the study. Teachers initially believed that a “good” teacher should focus on all the errors and provide high quality, comprehensive WCF. This standpoint reflected the “error hunter” (Hairston, 1986) mentality where teachers want to provide very comprehensive error corrections in students’ writing. However, all of the teachers in this study were surprised to find that students preferred indirect WCF over direct WCF. They agreed that indirect WCF was more effective in promoting acquisition of the target, but noted that it was time and energy consuming to provide indirect WCF with examples. One teacher mentioned the efficiency benefits in providing feedback through direct WCF, allowing her to put attention elsewhere. Chandler (2003) also found direct correction to be the fastest method, and indirect with metalinguistic codes and examples to be the slowest method. The perspectives of the language teachers were therefore mixed—there are apparent benefits to indirect WCF, especially on certain grammatical targets, but it will consume more of a teacher’s time. Adapting WCF research findings to pedagogical practice in real writing classrooms, while considering their real-world constraints, is still very difficult for teachers and researchers alike (Lee, 2018). More studies are needed to incorporate teachers’ perspectives on
WCF and consider practical methods for teachers to adopt WCF to enhance the effectiveness of their WCF practice.
CHAPTER VI: LIMITATIONS, CONCLUSION, AND FUTURE RESEARCH

6.1 Limitations

The current study has several limitations that should be considered when interpreting the results: First, it is impossible to directly assess the benefits of WCF methods compared to other instructional techniques due to the lack of a control group that received no WCF. In addition, there might be other variables from the coursework that contributed to the accuracy improvements. While such controls may be achievable in a controlled, laboratory setting, the current study was conducted in the authentic classroom context, so it was unethical to withhold all WCF from a group of students when research has clearly demonstrated WCF is beneficial to writing accuracy, fluency and complexity (e.g., Ferris, 2006; Shintani et al., 2014). The classroom setting also adds potential realism benefits that balance these limitations. SLA researchers have recently called for conducting research in classroom contexts to add ecological validity to the study and increase teachers’ usability of the findings (e.g., Leow, in press; Leow & Manchón, in press;). Lee (2019) advocated for examination of the effectiveness of WCF in authentic classroom contexts so teachers would understand the findings and apply them in their daily feedback practices. To limit data contamination, the researcher conducted a survey about target items exposure and extra learning outside that of the course itself and eliminated other variables that would be responsible for the targets (le and classifiers) accuracy improvement. By excluding any participants’ results that were contaminated by outside influences, the data can be reasonably isolated and primarily reflect the effects of the assigned feedback treatment.

Second, there is a risk that the data collection methods themselves affected the results. The concurrent think-aloud protocol is a type of introspective data elicitation procedure in which the participant describes their thought processes while performing the assigned task (Ericsson &
An advantage of think-aloud protocols lies in their ability to provide insights into the hidden cognitive processes that occur while performing a task, increasing the internal validity of research results (Leow, Grey, Marijuan, & Moorman, 2014). However, at the same time, think-aloud protocols are potentially intrusive procedures that have a potential for reactivity with performance of the task. In other words, it is possible that the act of thinking aloud may influence the cognitive process(es) under investigation or add an additional processing load on the participant. Studies that investigated think-aloud reactivity generally found that there is not a significant reactivity effect from thinking aloud (e.g., Leow & Morgan-Short, 2004; Polio & Wang, 2005; Bowles, 2010). The reactivity effect can be minimized through effective experimental design. In the current study, since both WCF treatment groups employed think-aloud protocols and both treatment groups might have reactivity, the equal treatment should not compromise the internal validity of the study. Due to the low n-size, the study was unable to add separate direct and indirect WCF groups without think-aloud protocols. Ideally, think alouds would only be required for a subset of students that would provide qualitative data, but given that only a portion of think aloud data is useful, the overall sample size would need to be large enough to allow for this separation. In future research, having established learner processes in these two types of WCF, it would be ideal to conduct a similar study with two similar groups without think-aloud protocols.

Third, the participants’ backgrounds might also affect generalizability of this study’s results. Participants in the current study were selected from a task-based Chinese synchronous computer-mediated communication program that helps prepare students to live and work in China as mid- to senior-level professionals in large, multinational organizations. These participants were self-selected, highly motivated, and most of them have a multilingual
background with past foreign language learning experiences. Some of them have learned multiple, distinct foreign languages. Therefore, their experience learning Chinese and interpreting feedback might be qualitatively different from the average learner because of the significant learning strategies they are likely to have acquired during their education. Future studies may wish to examine learners with other backgrounds and compare their learning outcomes and attitudes toward WCF. In addition, individual differences such as working memory, proficiency levels, and language analytic abilities could be added as variables to further untangle the effectiveness of different types of WCF for students of different educational backgrounds. Given that a significant amount of foreign language learning happens in the post-secondary environment, the current study’s results should still be useful in many contexts.

Fourth, this study was only single blind—students were unaware of the difference in the treatments or the grammatical targets. The teachers and data collectors, however, knew the WCF treatments and targets, creating the potential for their knowledge to influence the instruction or measurement. This potential bias was controlled by having each teacher apply both methods to different students, and averaging results across multiple teachers.

There is also potential that the qualitative data was influenced by the student-teacher relationship. Only the teachers were able to interact with the students in the program, providing the WCF treatments, while also conducting the individual qualitative data-collection interviews. It is possible that some participants might not have shared their honest responses with their teachers due to social pressure and the feeling of authority. Teachers as researchers may elicit a halo effect on the students wherein the classroom experience is more likely to be praised (Mackey & Gass, 2016).
However, there are also advantages. As Hyland (2016) commented, a beneficial feature in the writing research “is that it tends to favor data gathered in naturalistic classroom settings rather than controlled conditions” and “that there has been a strong preference for collecting data in authentic circumstances not specifically set up for the research, such as via classroom analyses of naturally occurring texts” (p.121). It was much easier to build trust between student participants and teacher participants, and to express candid responses with a familiar person. The writing tasks implemented in the study were also designed based upon the course curriculum and teachers’ meetings, meaning that they were directly relevant and not unnecessarily sterile. The teachers’ focus group was also effective in obtaining authentic classroom data and teachers’ perspectives on WCF practices because the teachers all knew each other and had a good working relationship. Their shared experience over the course of the study created an atmosphere for open conversation on both positive and negative experiences that would be unlikely to come out with an independent observer.

Further, while the think-aloud data provides observational views into student engagement, the relatively small n-size of this study required the feedback questionnaire and interviews to expose students to the difference in the feedback types and ask for their opinions on engagement. This design asked students to reflect on a treatment they did not receive, as a hypothetical, looking at only a limited example set. In a more ideal setting, there would be enough participants so that they could reflect only on the treatment that they received, answering purely about how engaged they felt with the feedback. Thankfully, this is only one of multiple views into engagement in this survey and is used primarily to supplement and explain the think-aloud data, and should still be valid for supplementary analytical purposes.
Lastly, only two extreme types of WCF were examined in the study: direct WCF and indirect WCF with metalinguistic information. This study provides a strong foundation for further research that may be able to include other WCF types and patterns to better improve the applicability for teachers’ WCF strategies. In the authentic classroom, teachers may employ a variety of WCF methods in different permutations. It is unknown how variations, mixes, and orders of WCF types may influence the results, and there may as yet be some ideal combination of feedback that would lead to superior performances. Future research should explore the combinations of different types of feedback, the sequence of different types of feedback, and other influential variables, such as proficiency level, L1, and working memory to understand the feedback engagement.

6.2 Conclusion

Drawing on the strengths of previous research, the current study investigated the effectiveness of unfocused WCF in a task-based Chinese foreign language curriculum. Thirty-eight L2 Chinese beginners were divided into the two written corrective conditions: direct and indirect with metalinguistic explanations. All participants wrote Writing Task 1 as the baseline, completed the Revision as the immediate posttest, and finished New Writing Task 2 after two weeks as a delayed posttest. All writing tasks were designed based on the host institution’s Chinese task-based curriculum. Besides measuring accuracy improvement on the target linguistic structures, particle le and classifiers, the study also employed think-aloud protocols to measure how participants processed the two types of WCF (direct and indirect with metalinguistic explanations). Quantitative analysis of accuracy scores, think-aloud data, and other qualitative data (think aloud data, interviews, questionnaires, and focus group discussion) were triangulated
to demonstrate students’ engagement with the direct feedback and indirect feedback with metalinguistic explanations.

Accuracy results on both particle *le* and classifiers demonstrated that the direct WCF group statistically outperformed the indirect metalinguistic explanation group on the Revision of Writing Task 1 with a medium effect size on *le* and a small effect size on classifiers. However, the indirect-with-metalinguistic-explanations group performed significantly better than the direct WCF group on New Writing Task 2, with a large effect size on *le* and a small effect size on classifiers. Another interesting accuracy result was while both WCF groups significantly improved after each written corrective treatment on the Revision of Writing Task 1 with a large effect, their performance also significantly decreased on New Writing Task 2 with a large effect. In addition, the immediate effects of receiving WCF only lasted for the indirect-with-metalinguistic-explanations group.

The qualitative think-aloud data revealed the internal processes of participants working with WCF. Participants who received unfocused direct WCF overall processed at low levels of depth of processing, which showed that participants noticed or ignored correction, and they usually did not understand the underlying rules of particle *le* and classifiers. On the contrary, participants who received indirect WCF with metalinguistic explanations overall processed primarily at a high level, which indicated that this group of participants went beyond noticing the corrections, but also understood the grammatical rules and figured out related corrections on their own.

Lastly, triangulating quantitative analysis of accuracy scores, think-aloud data, and feedback survey results, the participants who received direct WCF had low feedback engagement, since the direct feedback only drew learners’ attention to the target structure at the
level of noticing. The indirect WCF with metalinguistic explanations guided participants to a high level of engagement, because indirect WCF with metalinguistic explanations encouraged learners to process at a deeper level and frequently reached awareness at the level of understanding. Since the direct corrections were not provided, participants in the indirect-with-metalinguistic-explanations group needed to figure out the correction through a mental problem-solving activity. When participants moved beyond level of noticing and processed deeper, this kind of depth of processing had the potential for higher levels of awareness (including at the level of understanding), facilitating better accuracy performance, especially on long term gains (Leow, 2015, 2019).

6.3 Future Direction

The findings on Chinese writing accuracy provided evidence to support the beneficial effects of WCF on writing accuracy, showing different types of WCF had different benefits. For future researchers who want to explore the effectiveness of different types of WCF, there are three methodological suggestions. First, studies should be situated in curriculums where writing naturally occurs. To reflect the authentic classroom context and be able to measure accuracy improvements, the current study designed writing tasks based upon a task-based curriculum, replicated teachers’ provision of unfocused WCF, and focused on two target structures for accuracy analysis. These factors should make it easier for teachers to apply findings in real language classrooms. Second, considerable attention has been paid to accuracy analysis in general, and the focus on writing accuracy improvements limits the understanding of WCF delivery. For example, in order to better track accuracy improvements, many studies provide WCF on only one target structure in a laboratory setting. The accuracy improvements do not reveal the full picture of WCF effectiveness because it only demonstrates that certain types of
WCF can be delivered effectively and do not reveal the feedback being received cognitively. How students process WCF and how students engage with WCF are important keys to unlocking the benefits of different types of WCF. Future researchers could employ think aloud and eye tracking as process measures at the same time, while combining cognitive and social theoretical frameworks to analyze learners’ concurrent data about WCF. For teachers, this study highlights the opportunity to design new tasks that are not centered on oral production. Well-designed writing tasks in a synchronous computer-mediated context could take advantage of the on-screen visibility of written task performance and WCF. The direct WCF could present the positive linguistic evidence and lead to accuracy improvement in the immediate posttest. Indirect feedback with metalinguistic explanations could guide students to notice and understand new linguistic knowledge through the problem-solving activities suggested in this study’s observational data. Teachers can promote grammatical noticing and improvement by adding writing tasks in existing task-based curriculums. In terms of WCF practice in the authentic classroom, teachers employ a variety of WCF strategies to improve students’ grammatical development. The current study found that a vital aspect to improve students’ grammatical development is to make students notice the WCF, identify the gap between their prior knowledge and the target rules, and make them figure out the solution. Teachers can play a more active role by creating rich indirect WCF with metalinguistic explanations to encourage students to process feedback more deeply, to identify the gaps between their internal linguistic knowledge, to make hypothesis, and to experiment with their own error corrections. In this way, students may have a better understanding of teachers’ WCF and have high engagement.

Lastly, in the face of the ongoing covid-19 pandemic and the growing interest in online language learning formats, the current study should provide encouragement and reference points
to other work in the field of online language education and associated linguistics research. While this study was wholly conducted before the start of the pandemic, the teacher and students conducted all sessions virtually, normally across time zones and national borders, managing many of the challenges of virtual-only interaction that have now become a common part of the educational landscape. The success of this study and its host program may serve to illuminate the types of usability challenges researchers and teachers are likely to face in this virtual environment, and further emphasizing the importance of the rapport between the researcher, teachers, and student participants.

This study found that technical factors can often present challenges for the student. When applying research methods to the online context, usability tests for all of the study’s tasks and protocols may be important to certify that features and designs meet both the researcher’s internal standards and benchmarks, as well as those of the student’s environment. For example, it is important to understand through testing what kind of internet connection participants have and what equipment they use to ensure that these do not unduly influence the virtual classroom. If some students do not have sufficient internet bandwidth for high-quality video or low-latency interactions, one must consider how that may affect results and if it can be controlled.

Instructions and processes from the study must be simple enough to follow outside of planned interactions with the researcher or teacher (who may be in a different time zones or otherwise occupied). In this study, some students had questions about think-aloud protocols after the initial instructions, necessitating a dedicated online training module on how to complete it, featuring examples.

As many of us have found anecdotally over the past few months, virtual interaction results in the loss of many of the small, but important features of human communication, such as
eye contact and physical cues. This format requires its own type of socialization and, for a classroom, it is crucial to build a positive relationship between the researcher, teachers, and students that can encourage valid data and high engagement. By relying on students to complete their own data recordings, students must feel comfortable expressing their thoughts to the teacher and researcher without some of the formal trappings of the face-to-face research context that may help ensure a feeling of security in responses. In this study, the researcher and teachers created supplemental videos with the aim of further humanizing them and explaining the purpose of the study to the participants. These videos were placed on YouTube for the students to watch at their convenience and served to preserve scheduled instructional time.

In closing, I hope this study affirms the need for additional research in online contexts. It shows that there is a measurable effect for different pedagogical techniques applied to different linguistic targets in an online context. The steps taken here to overcome many of the challenges of online research were developed during the course of the study and should provide encouragement for researchers and practitioners to get creative in tackling this environment. I hope that these methods and data can provide support for other researchers, teachers, and students in these times.
APPENDICES

Appendix A: Participant Pre-semester Learning Needs Questionnaire

1. What name do you prefer to be called?
2. What is your gender?
3. What is your age?
4. Have you studied Mandarin Chinese before?
5. If you studied Mandarin Chinese before, what are your weaknesses when using this language?
6. Why are you taking this course?
7. What level of proficiency do you hope to achieve by the end of this course?
8. What do you need to do in this language by the end of the course?
9. What language(s) have you studied in the past, if any? When did you study and what proficiency level did you achieve (Please use ILR or ACTFL scale if you received official scores before)?
10. How do you like to be corrected?
11. Is there anything else you would like your teacher to know about you?
Appendix B: Online Task-based Curriculum Mandarin Chinese Express

COURSE PARTICIPATION

We recommend that you spend 6-8 hours per week independently studying the assigned course content. In addition, your teacher will schedule weekly 45-minute review sessions with you.

COMPLETION REQUIREMENTS

Attendance Status - In order to achieve an attendance status of “Complete” you must attend at least 80% of the weekly review sessions (11/14 weeks). Students who attend less than 80% of the weekly review sessions will receive a final attendance status of “Incomplete.”

Final Grade - Students will achieve a final grade of “Pass” if they are able to complete at least 80 task-based objective points, as verified by their Mentor during weekly review sessions. Students who complete less than 80 course objective points will receive a final grade of “Fail.”

FINAL PROJECT

In addition to meeting the attendance and course completion objectives, at the end of the semester each student will be required to submit an audio recording or a written essay that demonstrates their ability to use the language skills that they have acquired during the course. Mentors will provide additional guidance on what they expect these recordings or essays to include.

COURSE SCHEDULE

Week 1 Task 0, Background about Chinese language

- Understand Chinese Pinyin and the character writing system
- Recognize the initials and finals in the Pinyin system
• Recognize the four tones and neutral tone

• Understand Chinese radicals, basic strokes, and both simplified and traditional characters

Week 2 – 5 Task 1 Greetings and self-introduction

• Lesson 1  Be able to introduce yourself and understand Chinese name culture
• Lesson 2  Be able to introduce other people (family members and friends )
• Lesson 3  Be able to find out who someone is
• Lesson 4  Be able to greet someone and have a small talk at a party

Week 6 – 10 Task 2 Shopping

• Lesson 1  Be able to recognize the names for fruits, vegetables, meats, and other foods
• Lesson 2  Be able to buy various kinds of things in the supermarket
• Lesson 3  Be able to exchange or return household goods, clothes, and electronics
• Lesson 4  Be able to bargain at the Chinese market (local or night market)

Week 10 – 14 Task 3 Appointment

• Lesson1  Be able to talk about dates and the time
• Lesson 2  Be able to make an appointment
• Lesson 3  Be able to talk about different locations
• Lesson 4  Be able to take public transportations or use transportation apps

Submit Final Project that demonstrates mastery of above objectives
Online course content sample screenshot of Task 2 Lesson 1 shopping in the supermarket
Appendix C: Writing Task 1

Chinese Express Language Learning Task 1 Introduction and Shopping

Instruction

The purpose of the writing task is to review and practice grammar and vocabulary you’ve worked on so far. Don’t use a dictionary or your notes. Don’t worry about making errors, because we will use this first draft as a learning opportunity to practice self-correction. When you leave the course, you will develop a sharp linguistic sense and be able to learn on your own (we hope)!

You can use this piece of writing as a part of your final project. Please write as much as you can so we have enough material to work on. Pinyin with tones will be good, if you don’t know Chinese characters. If you are not sure about certain vocabulary or grammar, English will be fine too.

Please write a narrative story (250 words) and record your writing process by saying aloud what you are thinking in your mind. Please see the think aloud example in the supplementary materials.

Setting:

You met with your Chinese staff Nana for the first time. You did a nice self-introduction. You also told her that you studied Chinese for 2 years and you hope to continue Chinese learning after finished the visa work. You also told Nana about your previous work and your recent trip (you can decide where the trip was). You and your family ate nice food and visited interesting places during the trip. It was nice to know her and you guys went to the local market later.
In the market, Nana bought two bottles of soy sources sauce, three books, a box of strawberry and a bunch of bananas. You bought the following items (You can also buy other things): A whole chicken, two cucumbers, three tomatoes, one pound of broccoli.
Appendix D: Writing Task 2 (Delayed Posttest).

Chinese Express Language Learning Task 2 Family Introduction and Shopping Conversation

Instruction

The purpose of the writing task is to review and practice grammar and vocabulary you’ve worked on so far. Don’t use a dictionary or your notes. Don’t worry about making errors, because we will use this first draft as a learning opportunity to practice self-correction. When you leave the course, you will develop a sharp linguistic sense and be able to learn on your own (we hope)!

You can use this piece of writing as a part of your final project. Please write as much as you can so we have enough material to work on. Pinyin with tones will be good, if you don’t know Chinese characters. If you are not sure about certain vocabulary or grammar, English will be fine too.

Please write a narrative story (250 words)

Setting:

1. Your friend Nana’s family, Lee and you (or your family) decided to try some Chinese food together. It was your first time to meet Naha’s family. There are three family members. Nana’s husband is a manager in the trade company. Nana’s son is 6 years old and wants to visit US one day. You all ate five plates of dumplings (different flavors) and drank four glasses of beer. You also had a nice conversation with everyone (you can write anything you want to talk about, the more the better!)
2. After eating some Chinese food, you guys decided to go shopping in the mall and you brought some apples, a cellphone, a computer (You can also buy other things) and returned three books. Nana exchanged a bag and bought two shirts, and a pair of pants.
Appendix E: End-of-Course Interview and Survey

(both questionnaire and interview protocols were adopted from Bryfonski & Ma, 2019).

Teacher focus group interview

- Introductory questions
  - What was the best or easiest part of teaching the online course?
  - What was the most difficult part of teaching the online course?
- Methodology
  - In terms of having two distinct groups and feedback types, was it difficult for you at any point to stay consistent in your delivery of the written corrective feedback?
  - Did students ever force you to switch the feedback type you intended to use? E.g. by asking for more specific feedback in the direct written corrective feedback group?
- Written corrective feedback
  - From your perspective, what was the most effective form of feedback for students’ language learning? On tones, “le” and classifier? Why?
  - What was the easiest form of feedback to deliver from your perspective?
  - Between direct and metalinguistic feedback, what type did students seem to prefer or respond best to?
- Tones, “le” and classifier
  - From your perspective, which feature of the language was the easiest for learners to produce? Which was/were the most difficult?
- Results
  - What is your interpretation of the statistical results from the pre- and post-tests?
- Closing
  - Was there anything else you wanted to discuss that we didn’t talk about?

Learners: Mandarin Chinese Express End of Semester Interview

We hope you enjoyed your experience in the Distance Language Learning program. To assist us in assessing our program, we would like you to evaluate your training experience. Please highlight any strengths and weaknesses so that we can identify how we can improve the program for future students. Please note that teachers will not receive this feedback until after the semester has ended. Your responses will remain anonymous to the teachers.

1. What is your course?

2. What were your language learning goals at the beginning of this course? Did you achieve your goals by the end of this course?

3. Please tell us about a memorable time when you used the language skills you gained in this course effectively either on the job or in your personal life. What was the situation?
What did you do? What was the impact on section, mission, or Department goals? Even a few sentences would be extremely helpful to us.

4. Approximately how many hours per week in total did you spend on this course? Please include time spent studying and working with a teacher.

5. What do you think about the Course Design (e.g. course Content, presentation/layout, ease of navigation, instructions, pace of course)?

6. What do you think about your assignments (e.g. writing tasks, recording tasks, discussion board)?

Examples of direct and indirect feedback were shown by the teacher in the shared online workspace prior to subsequent questions.

†”Direct” feedback is provided by crossing out the error and inserting explicit corrections.

For example, a learner made a lexical error on the Chinese classifier bottle.

Chinese: 我买了一个瓶水。

Pinyin romanization: Wǒ mǎi le yī gé píng shuǐ.

Literal translation: I / buy / Particle le / one classifier (general) / classifier (bottle) / water

Translation: I bought a piece bottle of water.
In direct with metalinguistic explanation feedback, each error is marked and corrected by the teacher. “Indirect +” doesn’t give corrections but points out the error and provides metalinguistic explanations on how to correct the error.

Chinese: 我去 P-le1 学校。

Pinyin romanization: Wǒ qù P-le1 xuéxiào.

P-le1 means particle le usage 1. Particle le is used to express action completion, such as Wǒ chī le fàn (I ate dinner).

Literal translation: I / go / P-le1 / school.

Translation: I go to the school and verb “go” has the error code “P-le1”.

8. What kind of written corrective feedback did you receive on your writing tasks?

9. Did you appreciate the type of written corrective feedback you received or did you prefer another type? If so, please describe it.

10. What do you think about the written corrective feedback (e.g. timeliness and quality of feedback)? You may, of course, discuss any aspect of your training.

11. From your perspective, what was the most effective form of feedback for your language learning?

12. On tones, “le” and classifier learning, what were the easiest for you to produce? Which was/were the most difficult after receiving feedback?

13. Besides written correct feedback on “le” and “classifier”, did you receive feedback or instruction on these two linguistic items?

14. From the writing task 1 to the writing task 2, besides teacher’s feedback on “le” and classifiers, did you get explanations (on your own or ask native speakers) on these two linguistic items?

Learners: Written Correct Feedback Survey
This survey was distributed to participants via Google Forms
1. Your native language/languages?

2. Have you studied any foreign languages? If yes, please tell us the languages and the proficiency level.

3. Are you a heritage learner of Mandarin Chinese? (Heritage learner – learned Chinese at home with family)

4. Have you studied Chinese before? If yes, how long?

Direct written corrective feedback (example below)

“Direct” feedback is provided by crossing out the error and inserting explicit corrections.

For example, a learner made a lexical error on the Chinese classifier bottle.

Chinese: 我买了一个瓶水。

Pinyin romanization: Wǒ mǎi le yī gè shuǐ.

Literal translation: I / buy / Particle le / one classifier (general) / classifier (bottle) / water

Translation: I bought a piece bottle of water.

5. What do you think about direct written corrective feedback?

Indirect feedback with metalinguistic explanations
In direct with metalinguistic explanation feedback, each error is marked and corrected by the teacher. “Indirect +” doesn’t give corrections but points out the error and provides metalinguistic explanations on how to correct the error.

Chinese: 我去 P-le1 学校。

Pinyin romanization: Wǒ qù P-le1 xuéxiào.

P-le1 means particle le usage 1. Particle le is used to express action completion, such as Wǒ chī le fàn (I ate dinner).

Literal translation: I / go / P-le1 / school.

Translation: I go to the school and verb “go” has the error code “P-le1”.

6. What do you think about indirect written corrective feedback with metalinguistic explanations?

7. What kind of written corrective feedback do you prefer?

8. What kind of written corrective feedback did you receive on your written homework from your Chinese teachers?

9. From your perspective, what was the easiest form of feedback for your language learning? (easy to understand errors and easy to make revisions?)

10. From your perspective, what was the easiest form of feedback for your language learning (easy to understand errors and easy to make revisions)

11. From your perspective what was the most effective form of feedback for your language learning? (understand errors, learn new information, or improve in the future learning)

12. What emotional reaction did you have when you saw teachers’ written corrective feedback (overwhelming, excited, nervous, hopeless, or happy, or mixed feelings)

13. Is there anything else about written corrective feedback that you want to share with us?
REFERENCES


Press.


Mahwah, NJ: Lawrence Erlbaum.


https://doi.org/10.1016/j.jslw.2004.08.001


Manchón, R.M., Nicolás-Conesa, F., Cerezo, L. & Criado, R. (Forthcoming). L2 writers’ processing of written corrective feedback: Depth of processing via written languaging. In
W. Suzuki & N. Storch (Eds.), *Languaging in language learning and teaching*. Amsterdam: John Benjamins.


Radecki, P., & Swales, J. (1988). ESL student reaction to written comments on their written


Truscott, J. (1999). The case for “The case against grammar correction in L2 writing classes”:


*Psychological Science, 5*, 249–252.


Zamora, C. (2017). The secret is in the processing: A study of levels of explicit computerized feedback in heritage and L2 learners of Spanish. (Unpublished doctoral dissertation), Georgetown University, Washington, D.C.


