THE EFFECTS OF COMPREHENSIVE WRITTEN CORRECTIVE FEEDBACK
ON THE REVISION AND ACQUISITION OF SPECIFIC L2 FORMS

BY

ANTONIO B. PÉREZ-NÚÑEZ

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Spanish
with a concentration in Second Language Acquisition and Teacher Education
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Associate Professor Melissa Bowles, Chair
Professor Silvina Montrul
Assistant Professor Jill Jegerski
Professor Charlene Polio, Michigan State University
ABSTRACT

The role of written corrective feedback (WCF) as an instrument to facilitate Second Language Acquisition (SLA) has attracted much attention over the past fifteen years. However, the results from existing research on the effects of WCF (e.g., Truscott, 2007; Van Beuningen et al., 2012) present contradictory findings due to important design limitations. First, studies on WCF have typically employed a pretest-posttest design and analyzed one or two new pieces of writing, which is clearly not sufficient when comparing the effects of a WCF treatment, as some authors have pointed out (e.g., Bruton, 2009a; Storch & Wigglesworth, 2010) since there are many instances in which the errors found in a new text bear no relation to the errors previously corrected. Consequently, their results can only assume that the students used the language knowledge gained from the feedback on a previous text (i.e., pretest) and applied it to a subsequent text (i.e., posttest). Similarly, the available research has not yet provided specific evidence on how grammar correction by means of WCF affects specific L2 forms and whether error revision actually leads to accuracy development or L2 acquisition. Furthermore, given that WCF research has solely focused on second language (L2) and foreign language (FL) learners, there is no information on whether heritage language (HL) learners can benefit from this type of focus-on-form intervention and if so, to what extent the effects are different for this population of learners.

This study aims to fill these gaps by comparing for the first time the effects of error revision with and without comprehensive WCF (i.e., the most common type of error correction in writing classes, whereby multiple grammatical error types are corrected in each text) on the production, revision and accuracy development of specific linguistic features over time. Specifically, this investigation sought to answer the following questions: (1) Do students who do revision with and
without comprehensive WCF differ in terms of error revision and accuracy development of specific linguistic forms over time? (2) Are there any differences between L2 and HL learners in terms of error revision and the accuracy development of specific linguistic forms over time? Additionally, Truscott’s claim (1996, 2007) that WCF could cause learners to focus on grammatical correctness at the expense of written fluency and complexity was also examined by comparing the participants on various measures of written complexity and fluency at the beginning and at the end of the WCF treatment.

The participants were thirty-six learners (24 L2 and 12 HL) enrolled in a fifth-semester Spanish composition course at a public U.S. university. As part of the coursework, all learners completed a five-minute daily writing warm-up assignment three times per week at the beginning of each class session. The corpus for the study consisted of a total of 385 texts produced over the course of 11 sessions during a 4-week time span. The WCF group (N=18) received indirect WCF (i.e., all errors were marked by the instructor by underlining the whole word) on these assignments. In contrast, the revision group (N=18) received no form corrections (as was standard practice in the course). Both groups were given five minutes at the beginning of each class to revise their writing from the previous class. The changes in the error rate of four structures - (a) regular/canonical gender marking, (b) irregular/non-canonical gender agreement, (c) omission of definite articles in obligatory contexts, and (d) the correct use of the present subjunctive - were measured and analyzed to compare the effects of the treatments on both L2 and HL learners.

Overall, the WCF treatment led the students to make more accurate and varied morphosyntactic error revisions, whereas revision without WCF primarily led the students to make surface-level corrections (i.e., accent marks and spelling). The results also revealed that the
group that received WCF significantly increased their accuracy in the correct production of definite articles in obligatory contexts; however, no differences were found in the accuracy development of the other three linguistic forms (canonical and non-canonical gender marking and the present subjunctive). Finally, a comparison between the WCF and no WCF conditions at the beginning and at the end of the treatment showed that error correction had no measurable effect on either the complexity (i.e., verbal density and lexical richness) or the fluency (i.e., number of words per minute) of their writing.

This study makes various significant contributions to the WCF debate and the fields of SLA and L2 writing. The results and implications of this investigation provide the most exhaustive description of the factors that influence the efficacy of this teaching practice and explain them in terms of the acquisition process of each individual L2 form. It provides evidence of the limited effects of WCF and offers insight on how written production, error revision, and written feedback affect linguistic development at both group and individual levels by showing that for some forms, WCF can be effective not only to promote accurate revisions but also to increase learners’ written linguistic accuracy; however, it contextualizes the process in such a way as to show that WCF, as it is typically provided, is not a panacea that will fix any and all problems. Additionally, the data from L2 and HL learners show that differences in terms of proficiency and linguistic background may influence the extent to which these two populations could benefit from WCF and written production. Finally, the research methodology outlined in this study proposes a new framework that analyzes the role of WCF in production, revision and accuracy development, and also traces its effects on specific linguistic forms over time. This way, studies will not only be able to offer clearer evidence about the extent to which WCF can lead to accuracy development, but they will also serve to inform about and describe the true value of WCF in the SLA process.
ACKNOWLEDGEMENTS

I am deeply thankful to all of you who have made this possible. First, I would like to thank Dr. Inma Sanz from the University of Granada. She was an amazing mentor and support during my time in Spain and without her I would have never come to UIUC as an exchange student in 2009.

This dissertation was possible thanks to all the faculty, colleagues and friends in the Spanish and Portuguese Department who offered me emotional and intellectual support throughout all of these years, including those that I first met when I came as an exchange student from Granada that made my decision to pursue the Ph.D. at the University of Illinois at Urbana-Champaign an easy one.

I would also like to express my deepest gratitude to Prof. Melissa Bowles for her constant and exceptionally academic and professional support with this and all of my projects, and also to all the members of my dissertation committee, Dr. Montrul, Dr. Jegerski and Dr. Polio for their time, guidance and expertise. I would like to thank those who have been involved in this project; Christina De Angelo and Sarah West for their kindness by opening the doors of their classes and helping me conduct this study, and all of their students who agreed to participate; Florencia Henshaw for her invaluable advise and encouragement, Sara Fernández, and the undergraduate assistants, Laura Cummings and Gabriela Elizalde, for their effort and time aiding me. They all represent the outstanding people and companionship that has helped this department excel and that I have been very fortunate to experience throughout my years here.

My acknowledgement also goes to all my friends outside the department. Thank you to all of you in the GEO FC, Titivillus FC, United FC, and all the Intramural teams for all those exciting games, laughs and great moments. Thank you to my amazing tennis buddies, Fernando,
Alex and Carlo for their friendship and good times hitting balls and stress away. And thank you Benjamin ‘Malaguita’, Pocho and the rest of international grad students for all of the unforgettable times out and about in Champaign-Urbana.

I am also immensely thankful for my family and friends in Spain and across the world. Thank you Rafa, Juan Carlos, David, Otto, Nicky, and many others for your friendship in the distance and throughout the years. Thank you, Adolfo and Begoña, Tía Toñi, Antonio, Popis and Celicia for your support and for reminding me of my roots and who I am. Above all, I want to thank my mom, Dolores, and grandma, Maria, for inspiring me every day and for instilling in me the value for education, hard work and perseverance throughout my life.

Last and foremost, I would like to thank my loving Lauren for always being there and putting up with me during this process. You have been the most important part of all of this and there are no enough words to express my gratitude to you. Thank you for believing in me.
TABLE OF CONTENTS

CHAPTER 1: STATEMENT OF THE PROBLEM ................................................................. 1

1.1. Introduction ................................................................................................................. 1

CHAPTER 2: WRITTEN CORRECTIVE FEEDBACK AND SLA .................................. 8

2.1. Defining Written Corrective Feedback ................................................................. 8

2.2. Introduction ................................................................................................................ 9

2.3. Theoretical Foundations for the Use of WCF ....................................................... 10

2.3.1 Long’s focus-on-form ............................................................................................... 10

2.3.2. Schmidt’s Noticing Hypothesis .............................................................................. 12

2.3.3. Swain’s Output Hypothesis .................................................................................... 12

2.4. Theoretical Foundations against the Use of WCF ................................................ 15

2.4.1. Krashen’s Natural Order and Affective Filter Hypothesis ................................. 16

2.4.2. Pienemann’s Teachability and Learnability Hypotheses .................................. 18

2.4.3. Explicit versus Implicit Knowledge ....................................................................... 18

2.5. Types of Written Corrective Feedback ................................................................. 20

2.5.1. Introduction ............................................................................................................. 20

2.5.2. Direct and Indirect WCF ....................................................................................... 20

2.5.3. Metalinguistic WCF .............................................................................................. 22

2.5.4. Focused and Comprehensive WCF ...................................................................... 24

2.6. Research on the effectiveness of WCF in new pieces of writing .......................... 25

2.7. Corrective Feedback and Heritage Language (HL) learning ................................ 31

2.8. Summary of Relevant Research ............................................................................. 33
4.2.1. Research question 1a: Canonical Gender Marking ........................................ 67
4.2.2. Research question 1b: Non-Canonical Gender Assignment .......................... 78
4.2.3. Research question 1c: Definite Articles in Obligatory Contexts .................. 86
4.2.4. Research question 1d: Present of Subjunctive ....................................... 92

4.3. Research question 2: Comparison between L2 and HL learners ..................... 100
4.3.1. Research question 2a: Canonical Gender Marking .................................. 101
4.3.2. Research question 2b: Non-Canonical Gender Marking ........................... 114
4.3.3. Research question 2c: Definite Articles in Obligatory Contexts .................. 127
4.3.4. Research question 2b: Present Subjunctive .......................................... 134

4.4. Research question 3a: Complexity and Fluency Results ............................... 146
4.4.1 Research question 3a: Complexity Results ............................................. 146
4.4.2 Research question 3b: Fluency Results ................................................... 149

CHAPTER 5: DISCUSSION AND CONCLUSION ....................................................... 151

5.1. Introduction .................................................................................................... 151
5.2. WCF and No-WCF group error revision and accuracy development .................. 151
5.2.1. Canonical Gender Marking ...................................................................... 152
5.2.2. Non-Canonical Gender Marking .............................................................. 158
5.2.3. Definite Articles in Obligatory Contexts ................................................... 169
5.2.4. Present Subjunctive .................................................................................. 176

5.3. Comparison between L2 and HL learners within and across groups.................. 189
5.3.1 Canonical Gender Marking ....................................................................... 189
5.3.2. Non-Canonical Gender Marking .............................................................. 199
5.3.3 Definite Article Omission .......................................................................... 206
5.3.4 Present Subjunctive ................................................................................... 216
5.4. Complexity and Fluency .............................................................................................................. 230

5.4.1. Complexity .............................................................................................................................. 231

5.4.2. Fluency ...................................................................................................................................... 233

5.5. Implications .................................................................................................................................. 235

5.6. Limitations and Directions for Future Research ........................................................................ 243

5.7. Conclusions .................................................................................................................................. 245

REFERENCES ...................................................................................................................................... 247

APPENDIX A: SAMPLE OF THE TREATMENT .................................................................................. 268

APPENDIX B: LANGUAGE BACKGROUND QUESTIONNAIRE .......................................................... 270

APPENDIX C: LIST OF WRITING PROMPTS ..................................................................................... 273

APPENDIX D: PERCEPTION QUESTIONNAIRE .................................................................................. 275

APPENDIX E: INDIVIDUAL ACCURACY TRENDS ............................................................................. 292
CHAPTER 1: STATEMENT OF THE PROBLEM

1.1. Introduction

One of the most common pedagogical practices in second language (L2) classrooms is the provision of written corrective feedback (CF), which is seen by most instructors as part of their job and is also expected by their students. However, the role of written corrective feedback (WCF) as an instrument to facilitate second language acquisition (SLA) remains unclear, and, more specifically, the extent to which grammar correction can help learners notice their errors and whether that leads to greater accuracy in subsequent writing.

The starting point of this ongoing pedagogical and academic debate regarding the value of WCF began with Truscott’s (1996) claim that grammar correction was ineffective and harmful, and should therefore be abandoned. In response, and championing the case for grammar correction, Ferris (1999) argued that Truscott’s claims were too precipitous given the rapidly growing research evidence supporting the effectiveness of CF. According to Ferris (2002), students need additional, adjusted intervention from their instructors in order to compensate for their limitations and also to learn strategies to help them find, correct and prevent errors.

In spite of the large amount of research over the past two decades, some major problems and criticisms regarding the study of the effects of WCF can be identified, including (a) the inability to provide linguistic evidence that L2 learners apply the knowledge acquired from feedback on an earlier piece of writing to new writing tasks (e.g., Bitchener, 2012; Bruton, 2009a) and (b) the lack of WCF research framed within SLA theory (Polio, 2012a).

Although there is growing evidence of the relationship between WCF and accuracy improvement over time, as Bitchener and Ferris (2012) noted, the research base has so far been limited to testing the effectiveness of focused WCF (i.e., the learner receives error correction on
one or two error types) with certain linguistic categories (e.g., the use of English articles). One of the problems with this pedagogical approach is that as Xu (2009) noted, a focus on one or two grammatical categories may lead students to consciously monitor the use of those target features, while ignoring others. In this sense, focused WCF has been criticized for not taking into account one of the goals of language classroom instruction and the purpose of grammar correction, namely, to help students improve their accuracy in general, not one or two grammatical features (e.g., Ferris, 2010; Storch & Wigglesworth, 2010; Van Beuningen, et al., 2012).

Some studies, on the other hand, have investigated the effects of comprehensive CF, which is the most widely used grammar correction approach in language classrooms by which learners receive error correction on multiple error types at the same time. The studies (Hartshorn et al., 2010; Truscott & Hsu, 2008; Van Beuningen et al., 2008; Van Beuningen et al., 2012) that have investigated the efficacy of comprehensive WCF on new pieces of writing have assessed the outcome of their treatments by means of general measures of accuracy, including ratio/percentage of error-free words (Truscott & Hsu 2008), error-free clauses (Hartshorn et al., 2010), and error rate in broad categories such as ‘grammatical’ and ‘non-grammatical’ (Van Beuningen, et al., 2012).

The results and conclusions of these experimental studies assume that the students in the experimental groups draw (or do not draw) on language knowledge gained from the feedback on a previous writing task (i.e. pretest) to a second writing task (i.e., posttest). Thus, if no significant differences in the average error rates between the two groups (i.e., control and experimental) are found in a second text, it is then assumed that the students in the experimental group did not use the information from the corrections. However, as Bruton (2009b) argued, there are many instances in which errors in a subsequent piece of writing bear no relation to the errors
previously corrected, so the use of these metrics provides little evidence on the effects of correction in subsequent writing. Furthermore, the author contended that global measures of accuracy may also run the risk of obscuring instances in which learning has occurred.

This question is crucial because finding evidence of how WCF affects specific linguistic forms would provide insight to assess the efficacy and pedagogical value of a particular CF treatment, and would also address Truscott’s claims (1996, 2007) that no form of correction can be expected to help learner acquire morphosyntactic and lexical knowledge. In order to do so, WCF research needs to be guided by SLA theory because, as Polio (2012a) explains, some theories can be “invoked to address the effectiveness or lack thereof of error correction” (p. 376). Likewise, within the writing-to-learn perspective, in which writing is seen as a tool for language learning\(^1\), SLA-based studies on written grammar correction are key to gaining a better understanding of the role of writing and grammar in L2 development.

However, in contrast to what occurs with WCF investigations, the connection between theory and research has been a common practice in the study of oral CF for many years (e.g., Lyster & Ranta, 1997; Ellis, 2003). Only a few longitudinal studies (Evans, Hartshorn & Strong-Krause, 2011; Hartshorn et al., 2010) have framed their work on the effect of comprehensive WCF within a SLA approach, namely, by applying principles of skill acquisition theory to design a WCF treatment. According to the authors, feedback should reflect “what the individual learner needs most, as demonstrated by what the learner produces” and both the writing tasks and feedback should be “meaningful, timely, constant, and manageable” (Hartshorn et al., 2010, p. 87).

---

\(^1\) The rationale behind the writing-to-learn perspective (Manchón, 2011) is rooted in Cumming (1990) who argued that “composition writing might function broadly as a psycholinguistic output condition wherein learners analyze and consolidate second language knowledge that they previously (but not fully) acquired” (Cumming 1990, p. 483).
Furthermore, research work in the area of oral CF, especially within both the interactionist and sociocultural perspectives, has developed a series of constructs and engaged in a productive discussion that has allowed for more robust, empirically-based studies (see Lyster, Saito & Sato, 2013). Particularly, the concepts of noticing and uptake have attracted a lot of attention in oral CF research (e.g., Lyster & Ranta, 1997; Ellis, 2001) as they may foster L2 learners’ ability to restructure their knowledge and subsequent acquisition. ‘Noticing’ refers to Schmidt’s (1990, 1995, 2001) noticing hypothesis and denotes “the conscious perception of surface-level linguistic phenomena (Sachs & Polio, 2007, p. 71). ‘Uptake’ is a response to the teacher’s explicit or implicit feedback on a linguistic feature, and can be considered successful when a student uses that feature correctly or understands it (Ellis, 2001, p. 286). However, as Ellis (2001) notes, the correct use or understanding of L2 forms does not indicate that the feature has been acquired; instead, he argues that it is necessary to investigate whether the learner is able to produce the correct form on subsequent occasions. However, the WCF research that has to some extent examined these constructs is still very limited. These investigations (Qi & Lapkin, 2001; Sachs & Polio, 2007; Storch & Wigglesworth, 2010), as described in Chapter 2, only examined the processing of the WCF, specifically, noticing and uptake during text revision. Hence, future research on comprehensive CF should investigate the extent to which error noticing (e.g., error revision) lead to accuracy development in a larger number of pieces of writing that allow for the observation of accuracy changes over time.

In addition, while researchers have begun to compare how different types of instructional methodologies affect both L2 learners and heritage language (HL) learners, the available research on WCF has only investigated L2 and FL learners. Heritage learners are students who have been raised in a home where the heritage language was spoken, and as a result, they can
speak or understands the heritage language, and is to some degree bilingual in English and the heritage language (Valdés, 2001). In Spanish HL learners and L2 learners, studies have found both similarities and differences between these two groups, including not fully developed linguistic abilities and similar types of L1 transfer errors. With respect to the differences, HL learners have acquired the language in naturalistic settings and may have less understanding of grammatical concepts and also less experience with literacy skills in the heritage language than their FL counterparts (Carreira, 2007; Montrul et al., 2008). Given the lack of research with this distinctive and overlooked population, further investigation is clearly needed to explain and compare the effects of error revision and grammar correction on the acquisition of L2 forms in HL learners.

Finally, another important question in the WCF debate has to do with the extent to which WCF may affect other areas of language development, including fluency and complexity. Truscott (1996; 2004; 2007) have claimed that error correction could motivate learners to avoid the forms or structures that were corrected and thus lead to less sophisticated or complex writing. Also, the use of WCF has been hypothesized to hinder fluency (e.g., the number of words written in a specified amount of time) as learners may monitor their production more carefully in order to avoid making errors. Furthermore, according to the Trade-off Hypothesis (Skehan, 1998), L2 writers or speakers may focus, either consciously or subconsciously, on one of the complexity, accuracy and fluency (CAF) areas to the detriment of the others. Given the theoretical implications and the potential impact of attention to form and accuracy on other areas of writing development, many authors (e.g., Bitchener & Ferris, 2012; Polio, 2012b; Truscott, 2007) have argued for the importance of controlling for complexity and fluency in WCF research and have also call for more studies that investigate this important question.
All things considered, the goal of this dissertation is to address some of the major gaps previously outlined in several ways. This study goes beyond the use of general measures of accuracy by tracking for the first time the effects of a comprehensive WCF treatment on the accuracy development and retention of specific linguistic forms. To that end, the written production of a group of L2 learners and HL learners who received comprehensive WCF is compared to that of a group of L2 and HL learners who did not receive grammar correction. Additionally, this study, grounded in SLA theory and previous research in oral CF (e.g., Lyster & Ranta, 1997; Ellis, 2003), is unique and unanticipated in examining the impact of WCF on the production, revision, and acquisition of specific L2 forms.

All of the results and findings are discussed in the light of the most relevant approaches in SLA (e.g., Noticing Hypothesis, Output Hypothesis, Teachability and Learnability Hypothesis, Natural Order Hypothesis, etc.) that can be invoked to address the effectiveness of the treatments, or lack thereof.

After briefly describing some of the gaps and areas of WCF that required further attention in this first chapter and how this dissertation aims to address them, Chapter 2 first provides the theoretical framework that guides this study, followed by a typology of the most common WCF strategies and a critical review of the relevant WCF literature. The chapter concludes with a description of gaps and limitations in the previous research, which lead to the research questions and hypotheses that frame this dissertation.

Chapter 3 describes the methodology utilized to answer the research questions, including how the key constructs are operationalized as well as providing an overview of the research design and methods. Chapter 4 presents a description of the results for each research question, and, finally, Chapter 5 provides a discussion of the results in the context of previous corrective
feedback studies, along with the implications of this study in terms of SLA theory and language pedagogy.
CHAPTER 2: WRITTEN CORRECTIVE FEEDBACK AND SLA

2.1. Defining Written Corrective Feedback

The overarching goal of this dissertation is to examine the extent to which both error revision with and without the provision of comprehensive written corrective feedback (WCF) in second language (L2) learners’ writing may favor the acquisition and retention of L2 forms over time. Before discussing the role of this common pedagogical practice in L2 learning, it is necessary to specify what is understood by feedback, in general, and corrective feedback, in particular.

Feedback can consist of either positive or negative evidence, or a combination of both (Long, 1996; White, 1988). Positive evidence involves providing the learners with models of what is grammatical and acceptable in the target language (TL), whereas negative evidence, also known as negative feedback or corrective feedback (CF), involves the provision of information about what is unacceptable in the target language (TL). Within this view, errors (i.e., deviations from the norms of the TL) and corrective feedback are considered core components of language learning and teaching.

A further distinction can be made regarding the medium in which the CF is provided depending on whether it is delivered orally or in written format. Oral and written CF differ in a number of ways, which ultimately affect how the information is processed. First, some authors (e.g., Adams, 2003; Santos et al., 2010; Manchón, 2011) argue that written output and written corrective feedback (WCF) can be more beneficial for students as learners have greater processing time to compare their output with the corrections they received, which, in turn, may increase the possibility that they will notice the gaps in their IL. In addition, while oral CF is usually directed at individual learners, the correction that is provided is also available to the rest
of the class, which makes oral CF feedback more unclear and more likely to go unnoticed, when compared to written CF where the recipient of the feedback is always clear (Lyster, 2004; Sheen 2010). Lastly, opportunities for uptake, which are hypothesized to foster L2 acquisition, are generally more limited in oral CF due to time constraint and the online and interactive nature of oral communication, which is not typically the case for written production, and WCF.

2.2. Introduction

Despite the widespread use of WCF in the language classroom, and its theoretical support (e.g., Schmidt, 1990; Swain, 1985), there is still no consensus on how and if WCF can indeed help learners become more accurate writers in the L2 (e.g., Ferris, 1999; 2002; Truscott, 1996; 2007).

Several theories and approaches to SLA have been used to both advocate for and oppose the use of WCF. Many authors (e.g. Ellis, 2005; Long & Robinson, 1998; Norris & Ortega, 2000; Skehan & Foster, 2001) have argued that meaning-based approaches to L2 instruction are not sufficient and that some attention to linguistic form is necessary to obtain native-like proficiency in the L2. Some of these approaches include Long’s focus-on-form (Long 1991; 2000; Long & Robinson, 1998), Schmidt’s Noticing Hypothesis (Schmidt & Frota, 1986; Schmidt, 1990; 2001) and Swain’s Output Hypothesis (1985; 1995), which take a cognitivist perspective in which L2 acquisition occurs by means of interaction between the input, the cognitive system and the learners’ perceptual system. The basic theoretical argument for the use of WCF is that output alone is not sufficient and that some sort of attention to form is necessary to make learners aware of a mismatch or gap between their interlanguage (IL) and the TL grammar.
On the other hand, the theoretical principles against the use of CF are grounded in the Natural Order Hypothesis (Krashen, 1982) and the Learnability Hypotheses (Pienemann, 1989), which suggest that L2 learners acquire grammatical features in a relatively predefined order and not in sequences implemented by a teacher or course syllabus. As a result, linguistic features that are beyond a learner’s stage of development will not be teachable. Challengers of the use of error correction also argue that, if grammar correction or instruction yields any L2 knowledge at all, this emerging knowledge is only explicit in nature, and will lead to ‘pseudolearning’ or the superficial acquisition of linguistic forms (Truscott, 1996). The following sections will expand on the theoretical foundations commonly used to argue for and against the practice of WCF in L2 learning and teaching.

2.3. Theoretical Foundations for the Use of WCF

2.3.1 Long’s focus-on-form

The field of language teaching has experienced significant fluctuations over the decades in the support of what Long (2009) described as “interventionist” or synthetic approaches, and “laissez faire” or implicit approaches. According to Long, the “interventionist” approaches (e.g., grammar translation, audiolingulism, etc.) center the attention on the target language by presenting structures, notions, functions or lexical items one at a time, which might be later used in communication. On the other hand, the “laissez faire” approaches (e.g., the Natural Approach, immersion, the process syllabus) do the opposite by presenting samples of the L2 and helping students analyze the input and infer the linguistic rules and words. The goal of this type of implicit teaching is to let the language self-organize in the learners’ minds, respecting their internal syllabuses. However, research in immersion settings, especially in Canada, where learners received content-based instruction in the L2 has shown that while these “laissez faire” or
implicit methods where sufficient for comprehension and the development of fluency, L2 learners often lacked grammatical accuracy (e.g., Lightbown, 1991; Lightbown & Spada, 1990, White, 1988, 1991). For example, White (1988, 1991) found that L2 acquisition could not progress by means of positive evidence alone, because, in cases of ungrammaticality, such as the use of null-subjects in English or adverb placement, only negative feedback could help learners to be on the right track. The findings from this early research on immersion contexts suggest that if learners do not receive frequent CF to detect a mismatch between the output they produce and the TL norms, linguistic fossilization might occur, and that some form of grammatical instruction might be needed, particularly on those structures that significantly differ from the learners’ L1.

Similarly, Long (1991, 2000, 2009; Long & Robinson, 1998) has maintained that pure implicit learning is not sufficient, and has argued that what he called focus on form is a necessary component in language teaching and learning, especially, if the goal is to help adult L2 learners reach native-like ability. This understanding of CF as part of SLA and the limitations of input alone are further explained by Long’s (1996, 1998) interactionist hypothesis:

Environmental contributions to acquisition are mediated by selective attention and the learner’s developing L2 processing capacity, and that the resources are brought together most usefully, although not exclusively, during negotiation for meaning. Negative feedback obtained during negotiation work or elsewhere may be facilitated of SL development, at least for vocabulary, morphology, and language specific syntax and essential for learning certain specifiable L1-L2 contrasts. (Long, 1996 p. 414)

Hence, focus or attention to form is hypothesized to allow L2 learners to recognize a linguistic problem and identify it in the input. Consequently, focus-on-form interventions, such as grammar correction, be they oral or written, are seen as beneficial because of their potential to direct the learners’ attention and help them notice their errors.
2.3.2. Schmidt’s Noticing Hypothesis

One of the main theoretical motivations of the focus-on-form methodology is found in Schmidt’s Noticing Hypothesis (e.g., Schmidt, 1990; 2001). According to the Noticing Hypothesis, SLA does not happen subliminally, but requires conscious attention in order for input to become intake. In this sense, ‘noticing’ is seen as the intake of grammar that results from paying attention to the input, while ‘intake’ refers to “input which becomes part of the learning process” (Batstone, 1996, p. 273).

With respect to the function of attention to form within the Noticing Hypothesis, Schmidt (2001) gives it an important role as it can make learners aware of “a mismatch or gap between what they produce and what target language speakers produce” (p. 6). This concept is commonly known as noticing the gap (Schmidt & Frota, 1986) or cognitive comparison (Ellis, 1995). Hence, pedagogical practices, such as CF are expected to support the SLA process by triggering learners’ noticing of gaps between the TL norms and their IL, and thus, lead them to subsequently restructure their developing grammar.

2.3.3. Swain’s Output Hypothesis

The use of CF as an instrument to facilitate SLA also finds support in Swain’s (e.g., 1985, 1995) Output Hypothesis. According to Swain, output is essential because it ‘pushes’ learners to carry out deeper linguistic processing than in receptive skills such as listening and reading. Swain (1985, 1995) points out three specific functions of L2 output that can benefit SLA. First, the output that learners produce allows them to test hypotheses about the TL grammar, and help them move from a “purely semantic analysis of the language to a syntactic analysis of it” (Swain, 1985, p. 252). Moreover, both output, whether oral or written, can help learners engage in metalinguistic reflection that has positive effect on L2 learning, because as
learners reflect on their own TL, “their output serves a metalinguistic function, enabling them to control and internalize linguistic knowledge” (Swain, 1997, p. 119). Lastly, the author suggests that output has the ability to help learners notice ‘holes’ or gaps in their knowledge. Noticing of a gap activates, according to Swain and Lapkin (1995), mental processes that lead to the production of reprocessed or modified output, which may be evidence of “internalization of new linguistic knowledge, or the consolidation of existing knowledge” (p. 374).

However, in spite of the benefits of output, Swain (1993) also acknowledged its limitations in serving these three functions and stated the importance of CF as a complement by stating that:

“If students are given insufficient feedback or no feedback regarding the extent to which their messages have successfully (accurately, appropriately, and coherently) been conveyed, output may not serve these roles” (p. 98).

Furthermore, it is important to note that not all output is created equal, particularly with respect to modality: written or oral production. In this regard, some authors (e.g., Adams, 2003; Manchón, 2011; Williams, 2012) claim that engaging in L2 production through writing offers several advantages over oral output. According to Williams (2012), the value of output is stronger in written than in oral production mainly due to time constraints and the permanent record of the written medium. The author explains that, given that writing is slower than speaking, learners have the possibility of stopping the grapho-motoric process and concentrating on specific aspects of writing, including the retrieving of linguistic information. As a result, learners are not only more likely to visually notice their errors and gaps in their knowledge, but they also have more time to repair their communication problems by consulting other external sources or by simply reflecting on their explicit knowledge during the composing process. Also, the permanent and graphic nature of writing may increase the attention to formal language
features and facilitate hypotheses testing in the form of self-monitoring, reviewing, and evaluating their L2 production. Furthermore, Adams (2003) pointed to the advantage of written production, arguing that CF might produce a cognitive overload due to limited attentional resources and processing capacity during online production. Similar arguments are also found when comparing the value of written and oral CF as WCF typically provides learners with more time to process and compare their output with the corrections they received, which, in turn, may increase the possibility that they will notice the gaps in their IL (Sheen, 2010).

More recently, Williams (2012) has presented a model based on Housen and Pierrard (2005) that illustrates for the first time the role of both written production and CF in L2 development. This model consists of three steps that, as shown in Figure 2.1, are not completely independent, but influence one another and may also overlap during the process. As it can be observed, it incorporates some of the most relevant SLA notions, including the concepts of noticing, intake, output and retrieval.

Figure 2.1
*Processes involved in L2 writing and language development*

![Diagram](attachment:image.png)

(Williams, 2012, p. 322)

The initial step, known as ‘knowledge internalization’, is one of the components in the creation of new L2 knowledge and implies the noticing and processing of selected input. The second stage, ‘knowledge modification or restructuring’, refers to the learners modification or refining of their initial form-meaning connections. These two initial stages are, according to
Williams (2012), involved in the process of L2 knowledge creation. Focus on form, including CF, is essential during this process as the additional information and the learners’ response to it can help them confirm, reject and potentially modify their L2 knowledge. The last sequence of this model involves ‘knowledge consolidation’, which learners achieve through repeated retrieval and deeper processing, and results in faster access to their knowledge and the ability to use it in more contexts.

2.4. Theoretical Foundations against the Use of WCF

Truscott (1996) began a heated debate in the field of L2 writing after claiming that “grammar correction has no place in writing classes and should be abandoned” (p. 361). Truscott argued both practical and theoretical reasons to support his claim, which, ever since then, have been addressed to some extent when challenging (or backing) this position. The practical problems outlined by Truscott relate to the ability and willingness of instructors to give grammar feedback, and also to the disposition of the students who receive the corrections. According to him, there are a number of issues, including training, consistency and time to provide CF that affects its efficacy. Also, Truscott (1996, 1999) maintains that even if teachers make the effort to provide quality feedback, students still may find it difficult to understand, demoralizing, or may not be motivated to use it.

With respect to the theoretical problems related to WCF, after analyzing a series of studies (Kepner, 1991; Semke, 1984; Sheppard, 1992; among others), Truscott concluded that error correction, as it is typically practiced, does not take into account the gradual and complex process of acquiring the forms and structures of an L2. More specifically, his case is mainly based on Krashen’s Natural Order Hypothesis and Pienemann’s Processability and Teachability theories, as Truscott questioned the readiness of learners to process and acquire linguistic forms
and structures that lie outside of the natural sequence of acquisition. These two theoretical arguments, central to understanding the position against grammar correction, are further discussed in the next two sections.

2.4.1. Krashen’s Natural Order and Affective Filter Hypothesis

One of the main and most influential nativist theories in language teaching and acquisition is Krashen’s Monitor Model. This model forms the basis of the Natural Approach and consists of five main hypotheses (Krashen, 1982).

First, the Acquisition-Learning hypothesis states that there are two different ways for adults to develop competence in the second language, namely acquisition and learning. Acquisition is, according to Krashen, a subconscious process similar to that of children when they acquire their L1, and it is achieved by means of “implicit learning, informal learning, and natural learning” (p. 10). On the other hand, learning refers to conscious and formal knowledge of a L2, including “knowing the rules, being aware of them, and being able to talk about them […], known to most people as grammar” (Krashen, 1982, p. 10). Krashen’s distinction between acquisition and learning also predicts that error correction has little or no effect on acquisition, although it might be useful for conscious learning.

The second hypothesis proposed by Krashen, known as the Monitor hypothesis, explains the relationship between acquisition and learning. While acquisition is responsible for initiating an utterance and for fluency, learning functions as a Monitor or editor. The ‘monitor’ acts before we speak or write (planning) or after we produce (editing and correcting), and it only occurs when the learner has enough time to focus on form or correctness. A similar debate is currently being held on the role of corrective feedback in the development of implicit and explicit L2 knowledge, which will be later discussed in this section.
The third hypothesis is the Natural Order hypothesis, which states that the grammatical features of a language are acquired in a fixed and predictable order. This hypothesis is based on early research (e.g., Dulay & Burt, 1974; Fathman, 1975; Makino, 1980) that found that some grammatical structures seem to be acquired earlier than others, and that such order appears to be independent of the learners’ age, L1 background or the conditions of exposure. It is important to note that since the order of acquisition, according to Krashen, cannot be altered by explicit teaching of features that the learner is not ready to acquire, the author rejects grammatical sequencing in language program syllabi when the goal of language is acquisition, and not learning (in Krashen’s sense).

Krashen’s fourth hypothesis, the Input hypothesis suggests that, assuming the correctness of the natural order hypothesis, ‘comprehensible input’ is a necessary condition to move from one stage of acquisition to another, which is input that contains understandable structures that are “a little beyond” (Krashen, 1982, p. 21) the acquirer’s current level of competence. In this model of acquisition known as $i+1$, in terms of Krashen, ‘understanding’ is defined as understanding of meaning or content, not the form of the message, and it is carried out with the help of context and extra-linguistic information.

The last hypothesis, the Affective Filter hypothesis suggests that L2 language acquisition is influenced by affective factors, including anxiety, stress, self-confidence and motivation. According to Krashen (1982), acquirers vary with respect to the level of their Affective Filters, and those with lower filters will be more open to the input and, thus, the input that they are exposed to will more easily reach the language acquisition device.

With respect to the discussion regarding the role of CF in L2 acquisition, according to Krashen (1982), the most serious flaw in error correction is its effect on the Affective Filter, as
learners “will try to avoid mistakes, avoid difficult constructions, focus less on meaning and more on form” (p. 75). Furthermore, Krashen indicates that error correction leads to the Monitor being ‘overused’, in which L2 users are constantly checking their output with their conscious L2 knowledge, and, as a result, they cannot speak with fluency in the TL.

2.4.2. Pienemann’s Teachability and Learnability Hypotheses

Similar to Krashen’s Monitor Model, Pienemann’s (1985) Teachability hypothesis grew out of earlier research (e.g., Meisel, Clahsen & Pienemann, 1981) that studied the sequential L2 acquisition of learners who did not receive classroom instruction. Pienemann argued that the features that are beyond a learner’s stage of development are not teachable because “the acquisition process cannot be steered or modeled just according to the requirements of formal instruction” (Pienemann, 1989, p. 57). The author, who tested this hypothesis with learners of different ages and in different learning environments (e.g., Pienemann, 1989, 1998), also concluded that instruction was most effective when it reflected the stage just beyond the learners’ current stage of interlanguage. The implication for grammar instruction, which also echoes Krashen’s i + 1 notion in relation to input, is that it can only be effective when it is close to the point in which it could be acquired naturally. As such, opponents of grammar correction (e.g., Truscott, 1996) have argued that CF, as it is typically given, that is by means of correcting a wide range of linguistic forms at the same time, is deemed to be ineffective, as the input provided is usually beyond the learners’ level.

2.4.3. Explicit versus Implicit Knowledge

A central question in the debate on the usefulness of grammar correction refers to the type of knowledge that is yielded. Challengers of the use of error correction (e.g. Krashen, 1982; Truscott, 1996) maintain that, if grammar correction or instruction yields any L2 knowledge at
all, this emerging knowledge is only explicit in nature. Explicit or declarative knowledge, that is conscious grammatical knowledge, is typically contrasted with the unconscious and easily accessible knowledge or implicit knowledge (e.g. DeKeyser, 2003; Krashen, 1981). Therefore, since only implicit L2 knowledge enables learners to communicate spontaneously and fluently, the explicit knowledge that comes from grammar feedback or instruction is considered of very limited value in production. According to these non-interface position proponents, explicit knowledge will never become implicit, and because of this, the learners’ IL is unsusceptible to CF and its application may only lead to ‘pseudolearning’ or the superficial acquisition of linguistic forms (Trustcott, 1996).

In contrast to this view, many SLA researchers converge on the position that there is an interface connecting implicit and explicit knowledge bases (e.g. DeKeyser, 1998; Schmidt, 1990; Swain, 1985). For example, Skill Acquisition Theory (e.g. DeKeyser, 1998; 2003) proposes that explicit knowledge can gradually become part of the language used through output practice. Similarly, other authors (e.g., Ellis, 2010; Manchón, 2010; Polio, 2012a) have stressed the potential of language production and CF, especially in the written medium, to consolidate L2 learners’ linguistic knowledge and facilitate the development of accuracy and explicit knowledge. For example, Polio (2012a) highlights the importance of explicit knowledge since L2 learners are more likely to tap into both implicit and explicit knowledge during written production, in contrast to oral production, which demands more implicit knowledge due to time constraints. Therefore, since most L2 learners focus on explicit grammar rules at some point during the writing process, the use of WCF to increase explicit knowledge should not be dismissed, even though the ability to convert explicit knowledge into implicit knowledge is debated.
2.5. Types of Written Corrective Feedback

2.5.1. Introduction

The study of WCF has primarily centered on teacher feedback. This emphasis is mainly due to the idea that L2 learners cannot always provide each other with accurate and reliable grammatical input (Russell & Spada, 2006). This focus has also been influenced to some extent by early studies on CF and L2 learners’ perception (e.g., Chaudron, 1984; Hedgcock, & Lefkowitz, 1994; 1996), which indicated that students tend to have less positive attitudes towards peer CF and rely more on native speaker or teacher feedback.

It should be noted, however, that WCF is a complex construct due to the multiple options available for correcting students’ writing, as well as the different ways in which learners can respond to it. For this reason, one of the main concerns with the research on error correction in L2 writing is the extent to which different feedback methodologies may help learners develop their written accuracy.

With respect to the available options for written error correction, Ellis (2010) identified three main strategies - (1) direct, (2) indirect, and (3) metalinguistic WCF; and also distinguished, between focused WCF (in which only one or two different error types are corrected in the learners’ writing) and unfocused or comprehensive WCF (where multiple errors times are addressed at the same time).

2.5.2. Direct and Indirect WCF

Direct WCF is the provision of the correct form by the teacher to the student, and it is typically done in writing by crossing out the unnecessary/incorrect word, phrase or morpheme, inserting a missing word or morpheme, and writing the correct form above or near the erroneous form.
Example 1

*Caminando* (*caminar*) *es bueno para su salud*

‘Walking is good for your health’

Advocates of direct CF (e.g., Bitchener & Knoch, 2010; Chandler, 2003) argue that this type of correction enables learners to instantly internalize the correct form as provided by their teacher and offers the kind of explicit information that is needed for testing hypotheses about the target language. However, other authors (e.g., Ellis, 2009) indicate that one of the potential problems with direct WCF is that, while it could help students gain access to the correct form, it may not contribute to long-term learning, as it requires minimal processing on the part of the student.

Another widely used corrective feedback strategy is indirect CF, in which, instead of the teacher providing the target form, the error is marked (i.e., underlined, circled or highlighted) and the student is left to correct the problem that the feedback has drawn to their attention. Some authors (e.g., Bitchener & Knoch, 2008; Hyland & Hyland, 2006) believe that indirect feedback might be more beneficial than direct WCF as it requires a higher level of awareness and promotes problem solving and hypothesis-testing on the part of the learner, which some authors have suggested may play an important role in SLA (Leow, 1997, 2000; Robinson 1995) gains.

Example 2

*Caminando* *es bueno para su salud*

‘Walking is good for your health’

Several early studies on WCF (e.g., Fathman & Whalley, 1990; Ferris & Roberts, 2001) have claimed that indirect written CF has a positive effect on the development of L2 writing accuracy. In contrast, Robb et al. (1986), who investigated four types of feedback, including
direct and indirect CF, reported no significant differences between the four types. Another study (Chandler, 2003) compared direct CF and indirect CF. In this case, the indirect CF group was required to correct their errors whereas the direct CF group just received an indication of their errors. The author reported that the students who corrected their errors improved their accuracy from the first to the fifth piece of writing without decreasing their fluency in comparison to the direct CF group.

Nonetheless, no firm conclusions can be drawn from these studies, due to the lack of control groups (Chandler, 2003; Robb et al., 1986) or longitudinal measures (Fathman & Whalley, 1990; Ferris & Roberts, 2001). Still, some of these studies show that having students do something with the error corrections, such as revising a text, rewriting or studying corrections, is a crucial factor to help students develop the acquisition of linguistic forms.

2.5.3. Metalinguistic WCF

The third main type of CF is metalinguistic WCF, which involves providing learners with explicit commentary or information about the nature of the errors they have made. Ellis (2009) also further distinguished between two types of metalinguistic WCF, depending on whether the feedback is provided by means of (1) error codes or (2) metalinguistic explanations of the learner’s errors. Error codes typically consist of abbreviated labels or symbols that correspond to different kinds of errors, and they are usually placed near the error or in the margin. In the following example, the label “INF” is inserted to indicate the learner that the infinitive form, instead of the gerund, is required in that sentence in Spanish.

Example 3

Caminando [INF] es bueno para su salud

‘Walking is good for your health’
The second type, metalinguistic explanations, is the provision of brief grammatical descriptions about the type and/or nature of the errors. These descriptions, which may also include examples and grammar rules, are commonly placed at the end of the learner’s text. Many times, both direct and indirect WCF are accompanied with metalinguistic explanations in order to guide students in their revisions and give them grammatical information about what is wrong with their use of a linguistic form or structure.

Example 4

*Caminando* es bueno para su salud

Walking is good for your health

*Note: You need to use the infinitive when the verb is in subject position.

Finally, an important, and often overlooked, feature of WCF is the students’ response to the corrections provided. Ellis (2009) identified two types of responses, depending on whether the students are required to revise their errors or not. When revision is required, students may be asked to edit their errors and/or asked to study corrections. In contrast, when revision is not required, the students’ corrected texts are merely returned to them and, consequently, the corrections may be ignored. With respect to the importance of revision, Bitchener (2005) indicates that:

Requiring students to revise their writing in class immediately after they have received written feedback on their texts is one way of training students to become more independent and therefore more responsible for the linguistic quality of their writing. As Lalande (1982) and Lightbrown and Spada (1999) point out, such opportunities engage students at a *teachable* moment, that is, when they are working on their writing and are interested in the feedback they have received. (Bitchener, 2005, p. 3)
2.5.4. Focused and Comprehensive WCF

Another central question with respect to the use of error correction relates to whether WCF should be selective, or, on the contrary, address multiple errors at the same time. Selective or ‘focused’ WCF selects specific grammatical problems to be corrected and ignores other errors in that text. Ellis et al. (2008) also further differentiated, depending on the number of linguistic features selected, between highly focused CF, which emphasizes a single error type (e.g., errors in the use of the past simple tense) and less focused CF which targets more than one error type but still centers on a limited number of pre-selected types. Comprehensive WCF, on the other hand, involves correcting a wide range of errors in students’ written work. This type of feedback is also considered ‘extensive’ as it treats multiple errors at once.

Some researchers (e.g., Bitchener & Ferris, 2012; Ellis et al., 2008; Sheen, 2009) maintain that focused WCF might be more beneficial in terms of SLA than the comprehensive type. Sheen (2009) argues that the focused approach may enhance learning by helping learners to (a) notice their errors in their writing, (b) engage in hypothesis testing in a systematic way, and (c) monitor the accuracy of their writing by tapping into their existing explicit grammatical knowledge. On the contrary, comprehensive CF is, according to Sheen et al., (2009), more likely to (a) be given in a confusing and unsystematic way, and (b) become overburdening for the learner due to the attention to multiple error types. However, Bitchener and Ferris (2012) also acknowledge that providing feedback on only a few specific errors might neglect the learners’ long-term needs and affect their perception with respect to the errors they produce and their ability to edit their own texts comprehensively. Similarly, research also indicates that most students prefer error correction to be comprehensive, rather than selective or focused on limited
number of errors (Lee, 2005; Leki, 1991; Oladejo, 1993) in order to be aware of the different type of errors the make, especially when there is revision involved.

2.6. Research on the effectiveness of WCF in new pieces of writing

The role of grammar correction, and particularly that of WCF, as an instrument to facilitate SLA has attracted a great amount of attention from both researchers and language instructors for the past fifteen years. As was mentioned earlier in this chapter, the starting point of the debate regarding the value of WCF began with Truscott’s (1996) claim that grammar correction had no place in writing courses and should, therefore, be abandoned. In response, and championing the case for grammar correction, Ferris (1999) argued that Truscott’s claims were too precipitous given the rapidly growing research evidence supporting the effectiveness of CF. According to Ferris (1999; 2002), students need additional, adjusted intervention from their instructors in order to compensate for their limitations and to also learn strategies to help them find, correct and prevent errors. In the period following this debate, many researchers (e.g., Bitchener et al., 2005; Bitchener, 2008; Bitchener & Knoch, 2008; Chandler, 2003; Ferris & Roberts, 2001; Sheen 2007; Sheen et al., 2009) have claimed that CF can help improve grammatical accuracy. However, these studies have been criticized (e.g., Guènette, 2007; Van Beuningen, 2010) for their design limitations (i.e., the lack of a control group, focus on text revision rather than new pieces of writing, etc.) which have made it difficult to reach a firm conclusion on whether WCF indeed helps learners improve their grammatical accuracy over time.

Many authors (e.g., Bitchener, 2008; Bruton, 2009; Polio, 2012; Truscott, 2007; Van Beuningen et al., 2012) now agree that it is most relevant, from both a pedagogical and a theoretical perspective, to examine the effects of CF on new pieces of writing and that the analysis of the
errors that learners make or do not make over time can really inform us about the actual value of a particular feedback treatment in SLA, in contrast to the mere analysis of text correction, which can only be regarded as the initial step in the assessment of error correction.

Furthermore, based on what previous WCF reviewers and researchers (Bruton 2009a, Bruton 2009b; Bitchener & Ferris, 2012; Polio, 2012a; Polio, 2012b; Truscott, 2007; Van Beuningen, 2010; Van Beuningen et al., 2012) have said over the past years, a number of requirements can be now identified in order to design more controlled studies that can empirically test the effects of written grammar correction on SLA. These indications can be summarized as follows:

(a) There has to be at least one group that writes and is corrected on language form, which needs be compared to a group that writes, but is not corrected on language form (Bruton, 2009).

(b) During the treatment period, CF needs to be provided frequently in order to have any potential effect (Hartshorn et al., 2010).

(c) Students need to do something with the CF they receive (i.e., error revisions, rewriting the whole text or using error logs), in order to ensure that all learners pay attention to form (Bruton, 2009a; Ellis, 2010).

(d) The time on task (i.e., text revision, production, etc.) should be equal for all groups in order to set apart the effects of the WCF and make both the treatment and control groups comparable (Bitchener & Ferris, 2012; Van Beuningen et al 2012).

(e) The development of accuracy needs to be measured in multiple pieces of writing in order to be able to trace any changes in specific L2 forms (Bitchener & Ferris, 2012; Sachs & Polio, 2007).

(f) Changes in accuracy should be measured together with complexity and fluency, if
some form of development is intended to be measured (Bitchener & Ferris, 2012; Bruton, 2009; Truscott, 2007).

With respect to the most recent research designed to explore the effects of WCF on new pieces of writing, most of the available studies have centered on the extent to which focused CF can help learners improve linguistic accuracy over time (e.g., Bitchener & Knoch, 2008; 2010; Ellis et al., 2008; Sheen, 2007; 2010). The rationale for the use of focused CF, as opposed to comprehensive CF, is that correcting one or two recurrent error types at a time (e.g., use of definite and indefinite articles) might be more beneficial in helping the students notice and reflect on their corrections. The collective findings suggest that written CF works when it is intensive and concentrated on a specific linguistic error. However, these ‘focused’ WCF studies, which were conducted in laboratory settings, were criticized (Bruton, 2009; Xu, 2009) for their excessive focus on form and for narrowing their attention to grammatical accuracy and ignoring other aspects of writing, which bear little relation to current L2 classroom writing pedagogies.

On that account, several authors (e.g., Bruton, 2009; Van Beuningen, 2010; Xu, 2009) have called for more studies that investigate the learning potential of comprehensive WCF in the language classroom. This CF approach, which again involves the correction of all or most of the errors in the students’ writing, is usually considered the most pedagogically plausible type of WCF, in contrast to focused corrections (e.g., Lee, 2013; Williams 2012; Van Beuningen, 2012), since, as Van Beuningen et al., (2012) noted, one of the goals of correcting the students’ writing is to improve their accuracy in general, not one or two grammatical features.

Similarly, various researchers (Ellis et al., 2008; Hartshorn et al., 2010; Truscott & Hsu, 2008; Van Beuningen, De Jong, & Kuiken, 2008; 2012) have also stressed the importance of conducting such research in authentic classroom contexts in order to enable a valid assessment of
the role of CF on L2 learners’ written accuracy and learning. However, only a few studies have tested to some extent the effectiveness of the comprehensive approach in the language classroom. Two studies (Ellis et al., 2008; Sheen et al., 2009) compared the effect of focused and comprehensive WCF with L2 learners of English. Ellis et al. (2008) observed that, at the end of the treatment, the experimental groups were able to produce the target forms (English articles ‘a’ and ‘the’) more accurately than the control group that received no corrections and concluded that both focused and comprehensive WCF were equally effective. In a similar study, Sheen et al. (2009) found focused WCF to be more beneficial than comprehensive feedback when both approaches were compared to a control group. However, no information is given in these studies on what the participants did with the direct corrections that they received. This is of importance because, as it was indicated earlier, in order for any CF treatment to be useful, researchers and instructors need to ensure that students pay attention to and engage with the CF.

Another important point in question is that the efficacy of WCF, and especially that of the comprehensive approach strategies, cannot be measured by looking at only one or two forms. Recent studies have investigated the efficacy of comprehensive WCF (Evans et al., 2010; Hartshorn, et al., 2010; Van Beuningen et al., 2012) by employing a ratio that incorporates the sum of multiple error types. For example, Evans et al. (2010) and Hartshorn et al. (2010) examined a specific CF treatment for their ESL students that consisted of indirect coded CF using 20 error-correction symbols. Additionally, the experimental group received extensive error correction, including having to code, tally, and log all of their errors, as well as rewrite their 10-minute in-class essays until they were error-free. These short essays were written three to four times per week over a 13-week semester. The researchers reported that on the measure of accuracy (a ratio of error-free clauses to total clauses), the experimental group outsored the
control group. Although the overall results may indicate that students benefited from the WCF, the CF methodology requires a great amount of time and effort for both the instructor and the students. The treatment puts a heavy emphasis on grammatical accuracy including multiple revisions, rewriting of texts, coding and classroom discussions centered on the most frequent types of errors produced by the students. Some authors (e.g., Lee, 2013) have argued that the ultimate goal of WCF research should be to inform pedagogical practices, and consequently, CF studies need to implement practicable treatments that are in line with some of the current methodological trends and can ultimately be adopted in the classroom.

A third study, Van Beuningen et al. (2012) compared the effect of both indirect and direct written comprehensive feedback on the writing of L2 learners of Dutch. The authors found that both direct and indirect comprehensive CF led to improved accuracy in new pieces of writing (i.e., texts written during posttest and delayed posttest sessions, 1 and 4 weeks after the delivery of CF) when compared to two control groups (self-editing and mere writing practice with no CF). Unlike previous studies on WCF, Van Beuningen et al. (2012) is the first study to introduce a self-editing group in order to control for time-on-task, which is crucial in order to ensure that any improvements made by the experimental group are not due to the extra time spent on self-editing or revision.

With respect to their analysis, the authors provided a separate analysis of two categories of errors - ‘grammatical’ (i.e., a ratio calculated on the basis of the sum of the number of article errors, inflectional errors, word order errors, omissions of necessary elements, additions of non necessary elements, pronominal errors, and other grammatical errors) and ‘non-grammatical’ (i.e., lexical errors, orthographical errors, appropriateness/pragmatic errors, and other non-grammatical errors). According to Van Beuningen et al. (2012), only direct CF resulted in
grammatical accuracy gains in new writing, whereas the students’ non-grammatical accuracy benefited most from indirect CF. While the ‘grammatical’ and ‘non-grammatical’ error classifications may be seen as an advancement, as they go beyond the most commonly used general measures of accuracy (i.e., percentage of error-free words, error-free clauses, etc.), these categorizations are still too broad, as different error types are included in the same category. As a consequence, the results do not provide an actual indication that learning has occurred, nor do they provide any evidence that a learner is able to produce the correct form in subsequent production (Bruton, 2009a; Ellis, 2001). Furthermore, it would also be reasonable to believe that, as Ferris et al. (2000) found with their distinction of ‘treatable’ and ‘untreatable’ errors, some learners may be able to reduce both specific rule-governed error or ‘treatable errors’ (i.e., gender marking errors), as well as other ‘untreatable errors’ (i.e., lexical errors) that are supposedly less amenable to feedback.

Hence, as stated earlier, one of the basic requirements to assess the efficacy of WCF includes the evidence that a learner was able to provide the correct form in subsequent production after previously receiving (and attending to) the feedback on a specific form. This can only be achieved by means of a more detailed analysis that traces the impact of CF on the accuracy development of specific errors over time.

Furthermore, while noticing the gap between IL and L2 forms may facilitate the integration of L2 input contained in feedback into learners’ knowledge system, research in the field of oral CF suggests that ‘uptake’ (e.g., the correct revision of the target form or structure) might be predictive of subsequent L2 development (e.g., Egi, 2010; Loewen, 2005; McDonough, 2004; 2005; McDonough & Mackey, 2006; Shekary & Tahririan, 2006). Storch and Wigglesworth (2010), adopting a sociocultural theoretical perspective, attempted to investigate
whether uptake led to the retention of L2 forms in learners who received direct and indirect forms of comprehensive feedback. The study was conducted in an experimental setting and consisted of three sessions. In session 1, pairs of students composed a text following a prompt. In session 2, the pairs were asked to revise their errors and later rewrite the same texts. Finally, in session 3, each of the learners composed a text individually using the same prompt as in session 1. Pair-talk from session 1 and 2 was analyzed for instances of feedback processing, while correct changes in the revised text were considered evidence of uptake. As for retention, the authors were unable to provide evidence for it because the final and only texts individually written by the students “bore little resemblance to the text produced in pairs and contained new types of errors” (p. 320). The authors also acknowledged that a larger number of pieces of writing produced over time would have been necessary in order to trace and observe whether learners are able to repair their errors and to provide evidence that they use the correct linguistic forms in their subsequent production.

Given the lack of this type of research in the area of WCF, further investigation is clearly needed to explain the role of L2 written production and CF in terms of the relationship between error revision, uptake, and, subsequent production, in the process of SLA.

2.7. Corrective Feedback and Heritage Language (HL) learning

Many classrooms in the United States in which Spanish is taught enrolled together two distinct populations of language learners, which include the traditional foreign language learners who are monolingually-raised English speakers, and heritage language (HL) learners who were exposed to Spanish and English in childhood. Heritage learners are typically students who have been raised in a home where the heritage language was spoken and, to some extent, can speak or understand the heritage language (Valdés, 2001). Due to various factors (e.g., age, context of
acquisition, quality and quality of the input received, etc.) the linguistic competence of heritage speakers may range from very low proficiency, known as receptive bilinguals or ‘overhearers’, to highly proficient language users, who have both receptive and productive skills in the heritage language (Bowles, 2011). While research on HL and L2 learners of Spanish has shown that there exist similarities between these two groups, as both learners usually fail to develop full linguistic ability, including similar type of transfer errors from English and morphosyntactic issues (e.g., Montrul, 2005; Montrul & Bowles, 2009, 2010; Lynch, 2003), it remains unclear whether HL learners can benefit from the same type of instruction that L2 learners receive, given their differences terms of context of acquisition (e.g., natural environment vs. language classroom) and mode of acquisition (oral vs. written).

Some researchers have begun to compare how different types of instructional methodologies affect both HL learners and L2 learners of Spanish and their comparative effects on the acquisition of certain linguistic forms. These studies include processing and output-based instruction on the development of the Spanish past subjunctive (Potowski, Jegerski, & Morgan-Short, 2009), computer-based grammar instruction in the acquisition of the dative case marking in Spanish (Montrul & Bowles, 2010), the role of metalinguistic knowledge in the acquisition of the subjunctive (Correa, 2011). The first study of a classroom grammar treatment with HL learners (Potowski, Jegerski & Morgan-Short, 2009) compared the effects of both processing instruction and traditional output-based instruction in helping HL learners improve their accuracy in production and interpretation of the past subjunctive. The authors found higher accuracy gains with the use of processing instruction and suggested that processing instruction might be more beneficial than output-based instruction for HL learners and that they might benefit from focused grammar instruction. A second study (Montrul & Bowles, 2010)
investigated whether computer-based instruction in which the participants received immediate explicit feedback led to the accurate use of Spanish indirect object marker “a” with *gustar* type verbs in both elicited production and grammaticality judgment tasks. The authors found that instruction helped HL learners improve their accuracy and suggested a positive role of explicit instruction in the classroom. The last study, Correa (2011), compared L2 and HL learners in their ability to use explicit knowledge of terminology and understanding of rules (i.e., metalinguistic knowledge) of subjunctive in Spanish and their accurate production of such forms. The author found that L2 learners that were more aware of general grammar rules were able to use subjunctive more accurately than their L2 counterparts with less metalinguistic knowledge; however HL learners’ knowledge of grammar was not related to their accurate production of subjunctive. Correa (2011) concluded that HL learners might not need to learn grammatical rules in order to use the language. Considering the outgoing debate on whether the same type of instruction can benefit both L2 and HL learners, a crucial question is to understand the extent to which WCF can also help HL learners move forward in their interlanguage development in instructed settings.

2.8. *Summary of Relevant Research*

Many studies indicate that focused WCF (i.e., feedback that concentrates on one or two features at the same time) can favor the acquisition of certain linguistic forms (e.g., Bitchener & Knoch, 2008; 2010; Ellis et al., 2008; Sheen, 2007; 2010). However, relatively fewer studies have examined the efficacy of the most commonly used type of feedback, comprehensive WCF. The available research (e.g., Evans, Hartshorn, & Strong-Krause, 2010; Hartshorn, et al., 2010; Van Beuningen et al., 2012) that has investigated comprehensive WCF in new pieces of writing has assessed its efficacy by means of general measures of accuracy, which, as was argued before,
does not provide evidence on the role of comprehensive WCF to help learner acquire morphosyntactic and lexical knowledge and may obscure instances in which learning has occurred (Bitchener & Ferris, 2012; Bruton, 2009a; Truscott, 1996; 2007). Similarly, the available research on written grammar correction provides very limited information with respect to the SLA processes involved and the extent to which they influence L2 writing production and development of not only L2 learners, but also HL learners. In this regard, while noticing and uptake are hypothesized to favor SLA, no research to date has provided evidence as to whether error revision with and without WCF indeed helps the uptake of specific linguistic forms and whether it ultimately leads to subsequent accuracy development over time.

2.9. Research Questions

In light of the gaps in the literature on L2 writing and feedback for acquisition studies, in addition to the lack of empirical evidence of the effects of comprehensive WCF in the development of linguistic forms, the present study aims to address the following overarching research question: Are there any differences between text revision with and without comprehensive WCF on the accuracy development and retention of specific linguistic forms in learners of Spanish as a foreign language?

This question has been answered quantitatively and qualitatively by means of three types of data sources: (a) written texts produced by the learners before, during, and after the treatment and (b) a written attitudinal and perception questionnaire administered after the treatment. The analysis of the written texts has been used to answer the following questions:

RQ1. Do students who do revision with and without comprehensive WCF differ in terms of error revision and accuracy development of specific linguistic forms over time? Specifically:
a. Are there any differences in terms of error revisions and accurate production of canonical gender marking over time between the revision with and without WCF groups?

b. Are there any differences in terms of error revisions and accurate production of non-canonical/irregular gender marking over time between the revision with and without WCF groups?

c. Are there any differences in terms of error revisions and accurate production of definite articles in obligatory contexts over time between the revision with and without WCF groups?

d. Are there any differences in terms of error revisions and accurate production of present subjunctive over time between the revision with and without WCF groups?

RQ2. Are there any differences between L2 and HL learners in terms of error revision and the accuracy development of specific linguistic forms over time?

a. Are there any differences in terms of error revisions and accurate production of canonical gender marking over time between the L2 and HL learners in the revision with and without WCF groups?

b. Are there any differences in terms of error revisions and accurate production of non-canonical/irregular gender marking over time between the L2 and HL learners in the revision with and without WCF groups?

c. Are there any differences in terms of error revisions and accurate production of definite articles in obligatory contexts over time between the L2 and HL learners in the revision with and without WCF groups?
d. Are there any differences in terms of error revisions and accurate production of subjunctive mood over time between the L2 and HL learners in the revision with and without WCF groups?

RQ3. Does grammar correction have a negative effect on other areas of writing development?
   a. Are there any differences between the two group conditions in terms of complexity before and after the intervention?
   b. Are there any differences between the two group conditions in terms of fluency before and after the intervention?

2.9.1 Hypotheses

With respect to RQ1 and RQ2, given that WCF is hypothesized to promote error noticing, attention to form and guided learning (e.g., Bitchener, 2012; Van Beuningen et al., 2012), the participants who received grammar correction are expected to be more able to successfully identify those L2 forms that improved over time than students who did not receive correction. Furthermore, it is hypothesized that participants who receive comprehensive WCF will be more able to notice and modify their non target-like output, as shown by previous research on text revision (e.g., Ferris 2004; 2006). Moreover, if noticing (e.g., error correction) may serve as a predictor of L2 learning (e.g., Ellis, 1995; Robinson, 1995; Schmidt, 1990; Schmidt & Frota, 1986; Swain, 1985; 1995; Swain & Lapkin, 1995), then error revision of a linguistic feature will lead to intake of grammar knowledge, which, in turn, will result in more accurate production of those L2 forms over time.

The hypothesis for RQ3 is based on models of limited attention capacity (e.g., Skehan 1998; 2009) and previous research on L2 writing (e.g., Frantzen, 1995; Hartshorn et al., 2010; Sheppard, 1992; Truscott, 2007), which expects that a focus on accuracy due to WCF will
negatively affect other areas of L2 writing performance, including complexity and fluency, regardless of whether learners increase their accuracy or not. In terms of the WCF debate, Truscott (e.g., 1996, 2007) maintains that grammar correction lead to simplified writing, as L2 writers may avoid difficult constructions when writing a new text. Also, Krashen (1982) suggested that error correction might cause the Monitor to be ‘overused’, in which L2 users are constantly checking their output with their conscious L2 knowledge, which would result in no real fluency.
CHAPTER 3: METHODOLOGY

This chapter describes the methodology employed in order to answer the research questions that guided this investigation. First, a description of the participants’ characteristics and the classroom settings where the data was collected is provided. Then, the WCF feedback treatment and the revision procedures are explained, along with the instruments that were utilized in this study. Lastly, the coding procedures are discussed in detail, including the operationalization of the four linguistic categories (canonical gender agreement, non-canonical gender agreement, omission of definite article, and the use of present of subjunctive) that were investigated.

3.1. Participants

A total of 39 students enrolled in a sixth-semester Spanish composition course agreed to take part in the study. All of the participants were undergraduate students at a large public university in the United States enrolled in a sixth-semester Spanish composition course. They were all minoring or majoring in Spanish and their placement in the course was based on either their progression through the program curriculum or their score on a written proficiency test administered by the university. The data of three participants was not included in this study, as Spanish was their L3. The final dataset consisted of 36 learners of Spanish who were randomly assigned each of the two conditions; the WCF group (N=18) which received comprehensive grammar feedback on their errors, and the No-WCF group (N=18) that did not receive any type of form correction during the data collection period. Each group comprised 12 monolingually-raised native speakers of English and 6 heritage language learners of Spanish.

The most important characteristics of the L2 and HL learners in this study are summarized in Table 3.1 below. The results showed that the L2 and HL learners differed in
terms of first exposure to English and Spanish and self-reported proficiency in the two languages. The average self-reported age of exposure to Spanish in the L2 learners was 10.9 years of age in WCF group and 12.4 years of age in the No-WCF group, while all the HL learners reported being exposed to Spanish since birth. All the L2 learners were exposed to English since birth, while the average age of first exposure to English was similar between the HL learners in the WCF group (M=2.8) and their HL counterparts in the No-WCF group (M=3.0). The results in Table 3.1 also revealed that both the L2 and HL learners in the group conditions self-reported higher proficiency in English than Spanish. Additionally, the HL learners in the WCF group (M=4.33) and No-WCF group (M=4.67) reported higher proficiency in Spanish than their L2 counterparts (M=3.33), while all the learner groups reported to be more proficient in English, including the L2 learners in WCF and No-WCF groups (M=5), HL learners in WCF group (M=4.8) and the HL learners in No-WCF group (M=5).

<table>
<thead>
<tr>
<th>Table 3.1</th>
<th>Learner characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental Group (n= 18)</td>
</tr>
<tr>
<td>L2 learners (n=12)</td>
<td>HL learners (n=6)</td>
</tr>
<tr>
<td>Gender</td>
<td>9 females; 4 males</td>
</tr>
<tr>
<td>Age</td>
<td>Mean = 18.8</td>
</tr>
<tr>
<td>Range</td>
<td>18-20</td>
</tr>
<tr>
<td>Age of first exposure to Spanish</td>
<td>Mean = 10.9</td>
</tr>
<tr>
<td>Range</td>
<td>5-15</td>
</tr>
<tr>
<td>Age of first exposure to English</td>
<td>Birth</td>
</tr>
<tr>
<td>Range</td>
<td>0-6</td>
</tr>
</tbody>
</table>
### Table 3.1 (cont.)

<table>
<thead>
<tr>
<th></th>
<th>Initial Overall Accuracy (out 100)</th>
<th>Self-rated Spanish Proficiency (out 5)</th>
<th>Self-rated English Proficiency (out 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>89.8</td>
<td>3.33</td>
<td>5</td>
</tr>
<tr>
<td>SD</td>
<td>5.3</td>
<td>3-4</td>
<td>5-5</td>
</tr>
<tr>
<td>Mean</td>
<td>89.7</td>
<td>4.33</td>
<td>4</td>
</tr>
<tr>
<td>SD</td>
<td>6.7</td>
<td>4-5</td>
<td>5-5</td>
</tr>
<tr>
<td>Mean</td>
<td>86.1</td>
<td>3.33</td>
<td>5</td>
</tr>
<tr>
<td>SD</td>
<td>8.2</td>
<td>3-4</td>
<td>5-5</td>
</tr>
<tr>
<td>Mean</td>
<td>94.2</td>
<td>4.67</td>
<td>5</td>
</tr>
<tr>
<td>SD</td>
<td>3.7</td>
<td>4-5</td>
<td>5-5</td>
</tr>
</tbody>
</table>

3.1.1. L2 Learners

The L2 learners (n=26, 19 females, 7 males) were monolingually-raised native speakers of English (n=24), born and schooled in the U.S. They reported using only English at home with family and friends throughout their life. Their stays in Spanish speaking countries did not extend beyond one month, and none of the L2 learners received bilingual education in English-Spanish or were enrolled in dual immersion programs. Regarding their first exposure to the L2, 14 participants started studying Spanish in middle school, 6 in high school, and 4 took Spanish in elementary school. A mean comparison was computed to examine whether the initial accuracy scores (i.e., percentage of correct words) of the L2 learners who studied Spanish in elementary differed from their L2 counterparts who studied Spanish after elementary school. The results showed no significant difference between the L2 learners who were first exposed to Spanish in elementary school (M=90.8; SD= 6.3) and those who studied Spanish after elementary school (M=87.3; SD=7.1), $t(22)= 0.91, p = 0.37$. 
The initial overall accuracy scores of the L2 learners in the WCF and No-WCF, shown in Figure 3.1, were also compared in order to examine whether there existed any differences between the two groups before the treatment. The results revealed that there were no significant differences between the L2 learners in the WCF group (M=89.8; SD= 5.3) and the L2 learners in the No-WCF group (M=86.1; SD= 8.2).

Figure 3.1
Boxplot chart of initial accuracy scores by group and learner type: L2 and HL learners in experimental and control groups.

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; Overall Accuracy = percentage of correct words in first written text; L2= Second Language Learner; HL= Heritage Language Learner

3.1.2. HL Learners

The HL learners (N=12, 9 females, 3 males) were bilingually-raised Spanish/English speakers. 11 participants were born and schooled in the U.S. while one learner was born in Peru and moved to the U.S at the age of 8, completing the majority of schooling in English in the U.S. Nine of the twelve had both parents born in a Spanish-speaking country (8 from Mexico and 1
from Peru), two had one parent born in a Spanish-speaking country (1 from Mexico and 1 from Nicaragua), and 1 had two US-born Spanish-speaking parents. Most participants (N=9) indicated that their caregivers use Spanish when speaking to them, whereas the rest (N=3) reported the use of both English and Spanish. Half of the HL learners (N=6) reported that they ‘always’ used Spanish between the ages of 6-10; while 3 participants indicated that they ‘often’ used Spanish, and also 3 participants ‘seldom’ used Spanish. Similarly, more HL learners (N=5) reported that they ‘always’ used Spanish between the ages of 13-17, as compared to ‘often’ (N=4) and ‘seldom’ (N=3). In addition, all the HL learners had been exposed to English before the age of 8, and half of them (N=6) reported using both English and Spanish at home in childhood, while the other half (N=6) used only Spanish. The percentage of Spanish used by the HL learners at home was 51.7% (SD= 30.1), ranging from 5% to 100%. They all took Spanish language courses before entering the university and reported longer and/or more frequent stays in Spanish-speaking countries than their L2 counterparts.

The HL learners self-reported on average higher overall language ability in English (M=4.9) than in Spanish (M=4.5). A total of seven HL learners stated to be more proficient in Spanish than in English, four self-reported similar level of proficiency in English and Spanish, and only one HL learner indicated to be more proficient in Spanish than English. Similar to the L2 learners, a comparison between the initial accuracy scores of the HL learners in the group conditions before the beginning of the treatment revealed no significant difference between the HL learners in the WCF group (M=89.7, SD= 6.8) and their counterparts in the No-WCF group (M=94.1, SD= 3.6) groups, t(10)= -1.39, p = 0.193.

Lastly, a mean comparison of the initial accuracy scores of the L2 and HL learners was also computed in order to assess whether there existed differences between the two learner types.
before the treatment. The results indicated that the average initial accuracy score was higher in the HL learners (M=91; SD= 5.7) than in the L2 learners (M=88.1; SD= 7.0); however the differences were not statistically significant, \( t(34)=-1.63, p = 0.11 \)

### 3.2. Setting and Course Description

The data was collected from four intact classrooms from a sixth semester Spanish as a foreign language composition course taught by two different instructors at a public state U.S. university.

All four classes met for fifty minutes, three times a week during sixteen weeks. The main goals of this course were to provide many opportunities to practice writing in Spanish and prepare the students for the writing required in more content-based, upper-level Spanish courses.

The type of instruction in this course was mainly implicit and followed a learning-to-write approach, as learning happened in the form of reading and responding to good models, delivery of content-oriented feedback and the provision of holistic practice, in which providing form feedback was not standard practice. The course incorporated elements of a process approach to writing, involving activities such as free-writing, prewriting, the use of multiple drafts, revision and teacher and peer feedback (Ferris & Hedgcock, 2005). The instruction also included the exposure to and the teaching of a variety of genres, including description, narration and argumentation. More specifically, the course consisted of the following sections:

(a) Participation. Class participation and attendance constituted 10% of the final grade.

(b) Daily Writing Warm-up (*Ejercicios de Escritura Libre*). The first five minutes of every class period was spent writing about a specific topic provided by the instructor. The goal of these free-writing exercises was not to strive for grammatical or organizational precision, but instead, to explore various facets of the textbook chapter themes. These
exercises were graded for credit (1) or no credit (0), and consisted of 10% of the final grade.

(c) Grammar Exercises (Ejercicios de gramática). Students completed grammar exercises included in the first section of each chapter of the textbook. These exercises were graded credit (1) or no credit (0) and count for 10% of the final grade.

(d) Compositions. This final version of their composition consisted of a four-page paper (containing at least 1400 words). Students were required to submit a total of three essays, which consisted of four-page or 1400 words, throughout the semester. The first composition centered on description and narration, the second one on argumentation and the third on analytical writing, and counted for 40% of their final grade.

- Drafts (Borradores). Before submitting the final version of each of their compositions, the students were required to complete three drafts (borradores). The first draft consisted of a tentative thesis and at least four supporting points or main ideas. The second draft required an introduction, three well-elaborated paragraphs and a conclusion. The third draft should include all the suggested revisions from peer and teacher feedback sessions and four paragraphs. All drafts counted for 30% of the final grade.

3.3. Treatment

As briefly described above, as a normal part of the coursework, all students completed a five-minute daily writing warm-up assignment three times a week at the beginning of each class session. It was in this context that the current study was carried out. Specifically, the experimental group received WCF on their five-minute daily writing warm-up assignments three times a week over a 4-week period. In contrast, the control group received content but not
feedback on form, as it was the standard procedure for this in-class activity. For the purpose of this study, all writing and revision processes were completed online in individual writing pads that the students were granted access to. This online writing pad was used in order to let the students write, revise and automatically save their texts, as well as to allow the researcher to provide feedback faster and with more ease. Additionally, the writing pad did not include a spellchecker tool and no dictionaries or translators were allowed when producing their new texts.

The treatment consisted of comprehensive WCF in which different and multiple error types were corrected in the same text. The type of correction was indirect in that the correction was not directly provided; rather each error was marked in boldface. Similar to most studies on WCF, the error feedback was provided by the researcher conducting the study rather than by the classroom teacher in order to ensure that participants received similar feedback in terms of quantity, quality, and approach. According to Bitchener and Ferris (2012, p. 112) researcher feedback can help to control for some potential confounding variables, including (a) the ability and knowledge of the instructor to mark L2 forms and structures, (b) the amount of time, effort and attention devoted by each teacher and (c) other elements of the classroom, such as the structure and content of the lessons or the relationship between the instructors and the participants.

In addition to indirect correction, after marking the errors of each text, instructions and a summary with explicit information of the most recurrent error types was included at the bottom of each text in order to guide the students on their revisions, as in (1) and (2).

(1) “Please self-correct your errors. Check gender agreement and article omission and omission of "a". Add your correction in parenthesis next to the form in
bold face. You are free to use a dictionary or other online sources to help you revise. Gracias!"

(2) “Please self-correct your errors. Check indicative vs. subjunctive and gender and number agreement. Add your correction in parenthesis next to the form in bold face. You are free to use a dictionary or other online sources to help you revise. Gracias!”

This decision was pedagogically driven based on previous piloting of the WCF treatment in which the participants indicated that the indirect corrections (i.e., marking the error in boldface) did not provide enough information with respect to source of their errors and consequently were not able to correct them. The participants had access to their CF in the following class session in order to reduce the time between error production, corrective feedback and new written practice in which to potentially use the feedback (Hartshorn et al., 2010). Specifically, at the beginning of the next class session the students were given five minutes to look at their errors and correct them, before writing their new warm-up exercise/text on a new topic.

On the other hand, students in the control group did not receive any CF on their texts but were also asked to revise their writing during the following class session and also to type their revised forms in parenthesis next to the original form, following the indications below (3)

(3) Read through your text and revise it. Remember to add your correction in parenthesis next to the previous form. You are free to use a dictionary or other online sources to help you revise. Gracias!

Consequently, all of the participants had the same opportunity and amount of time allotted to revise their writing at the beginning of class in order to assure an equal distribution of
time-on-task across the two groups, and to make the feedback condition the only difference between the two groups. Lastly, as shown in the examples above, only during the time given for revision, the participants in the groups had the opportunity to use dictionaries/online resources and to ask the instructor for clarification. This decision was also made for both purposes of pedagogical authenticity and for the reason that accessing information from the Internet during revision is not, in many cases, dissimilar from asking an instructor a question. Furthermore, it should be noted that the access to the Internet for the purpose of revision was also intended to add ecological validity to the study given that it is a resource that most learner use when writing in their L2 and also in technology-enhanced language courses.

3.4. Instruments and Procedure

All the data included in this study were collected in two computer labs on the university campus over the course of four weeks in the fall semester of 2013. A summary of the research design and procedures is included in Figure 3.2 below.
Data collection was carefully scheduled to take place between weeks 6 and 9, as it was the time when argumentative writing, the genre selected for this study, was covered in class. Before that time, from weeks 1 to 5, the students worked on descriptive writing and narration. The data collection process comprised a total of eleven sessions, as shown in Figure 3.2. In session 1, participants completed the language background questionnaire and wrote their first text, which was used to establish the learners’ initial overall accuracy and also as the starting point in the production-revision sequence. The treatment period was implemented from sessions 2 through 9. There was no treatment during sessions 10 and 11 as to observe whether any accuracy changes were maintained over time and also to assure that the final measures for accuracy, complexity and fluency (session 11) were real rather than just an artifact of the proximity of the feedback.

In order to collect the data, four types of instruments were used: (a) a language background questionnaire, (b) writing tasks, (c) a keystroke logging tool, and (d) a perception questionnaire.

3.4.1. Language Background Questionnaire

In the first session, after signing the consent forms, the students that agreed to participate in the study filled out a written language background questionnaire (see Appendix B) adapted from Montrul et al. (2008). The questionnaire was designed to collect the students’ background data including age, gender, education, family background, years of Spanish study, and also inquired about the learners’ self-assessment of Spanish and English proficiencies. Based on their responses, the participants were categorized as either L2 or HL learners.
3.4.2. Writing Tasks

The writing tasks consisted of a total of 11 daily writing warm-up exercises (*Ejercicios de Escritura Libre*) that students were required to complete as part of their classwork. For each of their daily exercises, students responded to a prompt in Spanish, which was used to elicit the production of written argumentative texts (see Appendix C for a list of topics). The rationale for the use of argumentation, instead of spanning the entire semester and examining different genres, was to (a) control for the type of genre and (b) to elicit the production and allow for the comparability of genre-specific L2 forms and structures (i.e., production of subjunctive mood, omission of definite articles.) during the data collection period.

In addition, all the daily exercises were timed in order to foster unmonitored production intended to make learners more prone to tap into their implicit knowledge and thus allow for the measurement of linguistic development (Long, 2007). On the other hand, the revision sessions would let learners apply their explicit grammatical knowledge and observe whether the WCF and the No-WCF conditions differ in their ability to self-correct their errors.

All the writing warm-up exercises produced by the participants throughout the data collection period were analyzed in terms of the accuracy development of (a) regular/canonical gender marking (b) non-canonical gender marking, (c) definite articles in obligatory contexts (i.e., article omission) and (d) subjunctive mood. The basis for selecting these specific L2 categories is further explained in section 3.5.1.

Lastly, the first and last daily writing warm-up exercises were also analyzed to measure the participants’ written proficiency in terms of overall accuracy, fluency and complexity in order to (a) establish the proficiency level of the participants in both groups before the treatment
and (b) to evaluate the impact that the intervention had on various areas of writing development, including not only accuracy but also fluency and lexical and structural complexity as well.

3.5. Coding Procedure

3.5.1. Coding of Written Proficiency (CAF measures)

Following previous research on text quality and L2 development, a sample of the participants’ written production before and after the treatment was collected and coded for complexity, accuracy and fluency (CAF). The results aimed to offer information on (a) the initial overall accuracy scores of the participants (accuracy only) and also (b) to determine by comparing the participants’ first and last written text whether there existed any adverse effects on the students’ written fluency and complexity as a result of the provision of grammar feedback (Sheen, 2007; Truscott, 1996).

Similar to Unsworth (2008) the complexity dimension was evaluated by means of two ratio measures: (a) verbal density and (b) lexical richness. Verbal density was calculated by dividing the total number of finite and non-finite verbs by the number of T-units. Lexical richness was calculated by dividing the number of types by the square root of the number of tokens, also called Giraud’s Index. Only recognizable Spanish words were included in the count and words and names in English were excluded. This decision has been made because the use of such words (i.e., London, laptop, etc.) could otherwise artificially increase the scores of fluency and lexical richness. Accuracy in turn was measured by means of percentage of error-free words. The reason is that, as Polio (1997) noted, counting the number of errors, with respect to the total number of words, reports the quantity of errors better than other global measures of accuracy such as error-free t-units, which does not distinguish between one and multiple errors per t-unit. Also, given the proficiency level of the learners in this study, error-free t units would not have
been a very informative measure, since there were very few completely error-free t units in the
dataset, which could have led to floor effects. Lastly, fluency was measured as the total number
of words produced divided by the total time that a participant took to complete the writing.

3.5.2. Coding of Written Production and Description of the L2 categories

For the purpose of this study, four linguistic categories, (a) canonical gender marking, (b)
non-canonical/irregular gender marking, (c) article omission of definite articles in obligatory
contexts, and (d) present subjunctive, were selected in order to assess the effects of both revision
with and without WCF on the revision and accuracy development of specific L2 forms. These
specific linguistic forms were chosen based on the fact that (a) incorrect use of all those four
categories can be found in the production of L2 and HL learners of Spanish whose L1 is English,
regardless of their proficiency level, and (b) the forms vary in terms of how they are acquired
(i.e., rule-based as opposed to lexically stored) and in their degree complexity or difficulty (i.e.,
canonical gender marking vs. present subjunctive).

While previous experimental research has advanced on describing how these errors come
about, it is still unclear how these errors can best be overcome, if at all, and the role of both
written production and CF to foster the acquisition of these specific L2 forms and structures.

a) Canonical and Non-Canonical Gender Marking

Research has shown that both L2 and HL learners whose L1 does not have gender
marking (e.g. English speakers learning Romance languages) have great difficulty mastering the
use of gender with native-like ability (Franceschina, 2001; Grüter et al., 2012; Montrul et al.,
2008, 2013). Some authors (e.g., Carroll, 1989; Hawkins & Franceschina, 2004) claim that the
problem is due to maturational constraints and transfer effects, whereas others (e.g., White et al.,
2004) have indicated that gender feature is not subject to maturational constraints in L2 acquisition, and, as a result, transfer errors can eventually be overcome by L2 learners.

In Spanish, the grammatical category gender assignment is primarily morphologically and phonologically based. Most nouns follow formal rules; masculine nouns typically end in the vowel ‘o’ (e.g., *amigo* ‘friend’), and feminine nouns end in the vowel ‘a’ (e.g., *computadora* ‘computer’). However, there also exist other variants outside these regular patterns in Spanish. That is, some masculine nouns can end in the opposite vowels ‘a’, in the vowel ‘e’, or in a consonant, while certain feminine nouns can also end in the opposite vowels ‘o’, in the vowel ‘e’, or in a consonant. Similar to Montrul et al. (2013), in this study, masculine nouns ending in –o and feminine nouns ending in –a are referred to as canonical or transparent, and all other endings (-e, consonant, opposite vowels –a/-o) are labeled as non-canonical or non-transparent. Examples of these two types of noun endings are shown in Table 3.2, which were drawn from the texts the students produced.

**Table 3.2**

*Examples of Canonical and Non-canonical noun endings*

<table>
<thead>
<tr>
<th>Canonical</th>
<th>Non-canonical</th>
</tr>
</thead>
<tbody>
<tr>
<td>-o/-a</td>
<td>-e</td>
</tr>
<tr>
<td><strong>Masculine</strong></td>
<td><strong>Consonant</strong></td>
</tr>
<tr>
<td><em>amigo</em> ‘friend’</td>
<td><em>pie</em> ‘foot’</td>
</tr>
<tr>
<td><em>Feminine</em></td>
<td><em>computer</em> ‘computadora’</td>
</tr>
</tbody>
</table>

With respect to the source of gender marking errors, recent studies (e.g., Grüter et al. 2012; Montrul et al. 2013) have distinguished between assignment, agreement and ambiguous errors, as summarized in Table 3.3 below. In assignment errors, the learner has misclassified the gender of the head noun as in example (a). Agreement errors occur when the determiner and the
noun match but the adjective does not, as in (b). Ambiguous errors are those where the noun and
the adjective match, but the determiner and the adjective do not match, as shown in (c). They are
considered ambiguous since it is not clear whether the error is due to incorrect lexical assignment
or lack of agreement (Montrul et al., 2013).

Table 3.3

Types of gender marking errors

<table>
<thead>
<tr>
<th>Example</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. La casa blanca</td>
<td>Target</td>
</tr>
<tr>
<td></td>
<td>(‘the-fem. house-fem. white-fem.’)</td>
</tr>
<tr>
<td>b. El casa blanco</td>
<td>Assignment</td>
</tr>
<tr>
<td></td>
<td>(‘the-masc. house-fem. white-*masc.’)</td>
</tr>
<tr>
<td>c. La casa blanco</td>
<td>Agreement</td>
</tr>
<tr>
<td></td>
<td>(‘the-fem. house-fem. white-*masc.’)</td>
</tr>
<tr>
<td>d. El casa blanca</td>
<td>Ambiguous</td>
</tr>
<tr>
<td></td>
<td>(‘the-*masc. house-fem. white-fem.’)</td>
</tr>
</tbody>
</table>

The following coding scheme was applied in order to code the production of canonical
and non-canonical gender marking:

1. If a learner produced the phrase la casa (‘the-fem. house-fem.’), then it was coded as one
   instance of correct gender marking, as the noun (i.e., casa) was assigned the correct
gender (i.e., la).

2. If a learner produced the phrase *el casa (‘the-*masc. house-fem.’), then it was coded as
   one instance of incorrect gender assignment, since the gender assigned to the noun (i.e, 
casa) was incorrect (i.e., el).

3. If a learner produced the phrase *la casa blanca (‘the-fem. casa-fem. white-fem.’), then it
   was coded as one instance of correct gender agreement since the adjective (i.e., blanca)
   agreed with the noun (i.e., casa).
4. If a learner produced the phrase *la casa blanco* (‘the-fem. casa-fem. white-*masc.’), then it was coded as one instance of incorrect gender agreement since the adjective (i.e., blanca) did not agree with the noun (i.e., casa).

5. If a learner produced the phrase *el casa blanca* (‘the-*masc. casa-fem. white-fem.’), then it was coded as ambiguous since the determiner and the noun do not match (‘the-*masc. casa-fem’), while the adjective and the noun match (‘casa-fem. white-fem.’).

b) **Definite Articles in Obligatory Contexts**

The misuse or omission of definite articles in generic contexts in both oral and written production (e.g., Lipski, 1993; Montrul & Ionin, 2012) is another problem area common to both Spanish L2 and HL learners. With some exceptions (see Butt & Benjamin, 1988) that are outside the scope of this study, there is general agreement that the definite article is required before generic nouns in Spanish (e.g., *La gente estudia más idiomas hoy en día*), which differs from English, where the generic reference is expressed by means of bare plural noun phrases or NPs (e.g., ‘People study more languages nowadays’). The following list of generic nouns, extracted from Butt & Benjamin (1988), illustrates the contrast between Spanish and English:

1. Abstract nouns referring to a general concept, such as *la libertad* (‘freedom’), *la religión en México* (‘religion in Mexico’), *el debate sobre el aborto* (‘the debate on abortion’).

2. Countable nouns that refer to all the members of their class (e.g., *Los italianos comen mucha pasta*, ‘Italians eat a lot of pasta’).

3. Substances in general (e.g., *el agua es muy importante*, ‘water is very important’).

Consequently, the problem for L1-English L2 Spanish learners is that, since English has two forms (NP with definite determiner and bare NP) to convey two meanings (specific, generic) and Spanish has typically one form (NP with definite determiner) to express two meanings
(specific, generic), they tend to transfer the generic meaning to Spanish by omitting the definite article, as in Example 1a:

a. Delfines son inteligentes
   ⊗ [art omit]* dolphins are intelligent

   ‘Dolphins are intelligent’

With respect to the coding of this category, all the instances of use (and omission) of definite articles in generic contexts mentioned above were extracted and quantified. Then, the percentage of accuracy for this grammatical category (i.e., use of definite articles in generic contexts) was calculated for each text and the results were reported both individually and by group (see chapter 4).

c) **Present Subjunctive**

The contrast between the present subjunctive and indicative is one of the most difficult areas in the acquisition of Spanish, which affects Spanish L2 learners and Spanish HL learners of different levels (e.g., Montrul, 2002, 2004, 2009, 2011; Potowski, Jegerski & Morgan-Short, 2009). One of the main problems with the subjunctive mood for learners of Spanish is that, as Montrul and Perpiñán (2011) noted, it not only entails being able to produce correct morphology in oral or written production, but also knowing its syntactic distribution and comprehending associated semantic and pragmatic meanings. However, these meanings, according to the authors, cannot simply be accessed by observing contrasts between the forms and contexts, but rather are constructed from complex pragmatic inferences.

For purposes of consistency, only cases of present subjunctive in Spanish in the following contexts were included in the count and coded:
1. Present subjunctive that occurred in embedded contexts and were lexically selected by the verb (i.e., doubt/denial, volition and emotion), as in (1), or impersonal expression of the main clause (2):

(1) Quiero que hagamos /hacemos*
   I want that we do-subj/do-indic*
   ‘I want that we do’

(2) Es posible que el problema sea/es*
   It is possible the problem is-subjunctive/es*

2. Present subjunctive determined by pragmatic principles, including negation (3) and presupposition with relative clauses (4).

(3) No creo que sea/es justo
   I don’t think that it is-subj/is-indic* fair
   ‘I don’t think it’s fair’

(4) Necesitamos personas que trabajen/trabajan
   We need people that work-subj/work-indic*
   ‘We need people that work’

3. Oversuppliance of subjunctive by extending its use to contexts that require the indicative, as in (5).

(5) Creo que la sanidad sea muy importante
   I think that health care is-indic/work-subj* very important
   ‘I think health care is very important’
The occurrence of present subjunctive forms for each context and the total number of errors and correct production for the two conditions were calculated and reported both individually and by group in Chapter 4.

3.5.3. Coding of Revision

In this study, revision refers to the student’s attention to a particular error and the self-correction of such error during the revision sessions. The following coding scheme was applied:

1. If a learner successfully corrected an erroneous form during the revision session, whether indicated with WCF or not, it was coded as ‘correct revision’.
2. If a learner failed to correct an erroneous form during the revision session, it was coded as ‘incorrect revision error’.
3. If a learner did not attempt to correct an erroneous form during the revision session.

The total number and percentage of successful self-corrections for each error type was calculated in order to answer RQ1 and RQ2, that is, a comparison of the learners’ ability to revise their errors with and without the provision of WCF.

3.5.4. Coding of Accuracy Development

The percentage of correct use of a particular form, a measure that is consistent with the process of interlanguage (IL) development, helped answer RQ1 and RQ2, revealing whether the participants in the WCF and No-WCF differ in terms of accuracy development. The total number and percentage of correct forms was calculated for both gender marking and the definite articles in obligatory context and individual accuracy linear trends for each type were also computed in order to provide indication of accuracy development both individually and at group level.
3.5.5. Coding of Retention

Retention in this study refers to whether the revision of a particular form led to the correct use of the same form in subsequent production. Retention scores were calculated only for non-canonical gender marking and the use of present subjunctive forms due to methodological reasons. First, unlike canonical gender marking, there are multiple irregular forms or non-canonical ending nouns, hence calculating the accurate production of all the ‘non-canonical’ forms would not provide clear indication of the effect of revision, and may hide instances in which the correct or incorrect use of a specific form was produced. In the same vein, it would be counterintuitive to calculate the accurate production of all subjunctive forms given the variety of structures and differentiating contexts that elicit the use of this verb form. Consequently, a coding system that tracks both revision and subsequent production of specific forms would provide the required evidence to observe the extent to which revised forms are retained over time. For this purpose, the following scheme that comprised a total of six factors was applied to code the retention of revised forms in new pieces of writing:

1. Revision + Immediate Production: If a learner, after successfully correcting an error during the revision session, as indicated by the new form added in parenthesis, was able to produce the correct form in the subsequent new text, it was coded as 1; otherwise, if a learner, after successfully correcting an error during revision, was not able to produce the correct form in the following text, it was coded as 0.

In Example 2 below, a participant from the WCF group was able to successfully correct an initial non-canonical gender marking error ‘el mismos oportunidades’ by adding the correct non-canonical gender forms ‘las mismas’ in parenthesis during the revision session, as shown in
and later produced the correct non-canonical gender marking forms in the following text, as in (2). As a result, this sequence was coded as 1.

Example 2

(1) tienen el mismos (las mismas) oportunidades
   ‘they have the [masc]* same [masc]* (the[fem] same [fem]) opportunities
   [fem]’.

(2) abre la puerta a muchas oportunidades[…]
   Las oportunidades incluyen
   ‘it opens the door to many [fem] opportunities [fem] […]
   The [fem] opportunities [fem] include’

2. Revision + Later production: If a learner, after successfully revising an error during the revision session, produced the correct form in a later but not immediate text, it was coded as 1, otherwise 0.

   In example 3 below, the participant corrected an initial error ‘son’ (3rd person ‘be’, present ind.) by providing the correct revised verb form of the subjunctive ‘sean’ (3rd person plural ‘be’, present subj.) during the revision session (3). Given that the learner failed to provide the correct subjunctive form in a later text (4) that was elicited by the same main clause ‘no pienso que’ (‘I don’t think that’), this particular instance was coded as 0:

Example 3

(3) no pienso que son (sean) justos (Text 2)
   ‘I don’t think they are [PRESENT TENSE] fair’

(4) no pienso que el dinero da la felicidad (Text 6)
   ‘I don’t think money gives [PRESENT TENSE] happiness’.
3. Revision + Production in two different texts: If a learner, after successfully correcting an error during revision, produced the correct L2 form in two different texts, the sequence was coded as 1, otherwise 0.

Example 4 below illustrates how a learner, after successfully revising an initial non-canonical gender assignment error (5), used the correct gender assignment form in subsequent production (6), but was unable to produce the correct gender agreement in a second text (7). Consequently, this sequence was coded as 0.

Example 4

(5) Hay una (un) problema con  \hspace{1cm} \text{(Session 8)}
   ‘There is a [\textit{fem}]\* problema [\textit{masc}]’

(6) es un problema grande  \hspace{1cm} \text{(Session 9)}
   ‘it is a [\textit{masc}] big problem [\textit{masc}]’

(7) pero las problemas  \hspace{1cm} \text{(Session 10)}
   ‘but the [\textit{fem}]\* problems [\textit{masc}]’

4. No revision + Immediate Production: If a learner, after not correcting an error during the revision session, produced the correct form in a subsequent text, it was coded as 1, otherwise 0.

In the excerpt in example 5, the learner was not able to provide the correct non-canonical gender agreement form during revision (8), but produced the correct form in the following text (9). As a result, this instance was coded as 1:

Example 5

(8) a causa de costumbres nuevos  \hspace{1cm} \text{(Session 9)}
   ‘because of new [\textit{masc}]\* customs [\textit{fem}]’
5. No revision + Later production: If a learner, after not revising an error during the revision session, produced the correct form in a later but not immediate text, it was coded as 1, otherwise 0.

Example 6 shows a participant who did not revise a non-canonical gender agreement assignment during the revision session (10) and made the same error type in later, but not immediate, production (11). The following sequence was therefore coded as 0:

Example 6

(10) Nuestros educación (Session 2)

‘Our [masc]* education [fem]’

(11) No menciona el educación (Session 6)

‘It does not mention the [masc]* education [fem]’

6. No revision + Production in two different texts: If a learner, after not revising an error during revision session, produced the correct L2 form in two different texts, the sequence was coded as 1, otherwise 0.

In the example 7 below, the learner did not correct a non-canonical gender error during revision (j), and later produced the same error in two different texts, (12) and (13). This sample was coded as 0:

Example 7

(12) no es una problema (Session 3)

‘it is not a [fem]* problem [masc] customs’

(13) una problema en (Session 6)

‘it is not a [fem]* problem [masc] customs’
The researcher coded all of the data, including the production, revision and correct use of the four linguistic forms examined in this study, canonical gender marking, non-canonical gender marking, definite articles in obligatory contexts and the present subjunctive. Additionally, the written production of four participants (11.1%) was randomly selected and independently coded by another rater who was a native speaker of Spanish with college-level experienced teaching Spanish as a foreign language. All instances of agreement and disagreement between raters were counted and the Cohen’s Kappa coefficient of agreement between raters was computed for all categories.

Table 3.4
*Inter-rater agreement Cohen’s Kappa coefficient by category*

<table>
<thead>
<tr>
<th>Category</th>
<th>Total Production</th>
<th>Errors</th>
<th>Correct Revisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canonical Gender</td>
<td>0.87</td>
<td>0.80</td>
<td>0.88</td>
</tr>
<tr>
<td>Assignment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canonical Gender</td>
<td>0.73</td>
<td>0.648</td>
<td>0.75</td>
</tr>
<tr>
<td>Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Canonical</td>
<td>0.79</td>
<td>0.67</td>
<td>0.64</td>
</tr>
<tr>
<td>Gender Assignment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Canonical</td>
<td>0.77</td>
<td>0.83</td>
<td>0.65</td>
</tr>
<tr>
<td>Gender Agreement</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.4 (cont.)

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite Articles</td>
<td>0.55</td>
<td>0.59</td>
<td>0.82</td>
</tr>
<tr>
<td>Present Subjunctive</td>
<td>0.76</td>
<td>0.94</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Notes.* p < 0.001

The results showed that the level of interrater agreement ranged from moderate (0.55) to outstanding agreement (0.81-1.00) and were statistically significant for all categories, p < 0.001. Most statisticians agree that Kappa values should be at least 0.6 or higher to claim a good level of agreement (Landis & Koch, 1977). As such, it is possible to conclude that overall the level of agreement between the two raters was substantial (0.61-0.80). All cases of disagreement were discussed between the two raters until 100% of agreement was reached. It was found during the process of rater socialization that the variation in terms of coding was primarily due to some forms being unnoticed, especially in the case of article omission, or misclassified (i.e., gender assignment versus gender agreement). As a result, it is possible that additional training and/or the repetition of the inter-rater coding process would lead to higher level of agreement between the parties.
CHAPTER 4: DATA ANALYSIS AND RESULTS

4.1. Overview

As described in the previous chapter, the research questions were addressed by analyzing the participants’ written texts and their responses to the end-of-the-treatment perception questionnaire. The data were analyzed by means of t-tests, Chi squares, and correlations using SPSS 20 with all of the alpha levels for the statistical analyses set at 0.05.

The written dataset consisted of a total of 385 texts individually produced by 36 participants over the course of 11 sessions during a 4-week time span. The participants in the WCF group, which was comprised of twelve L2 learners and six HL learners, wrote a total of 192 texts, with an average of 95.54 words per text ($SD= 22.65$) whereas the participants in the No-WCF group, which also included twelve L2 learners and six HL learners, produced a total of 193 texts with an average of 92.26 words per text ($SD= 25.04$). The variation in the total number of texts between the two groups was due to one participant who missed one class session during the data collection period. Differences in terms of text length between the WCF and No-WCF groups were not statistically significant, $t(383)= 1.34, p = 0.180$; however, a comparison between the L2 and HL learners in the two groups revealed that the average text length, as indicated by the total number of words, was significantly shorter in the L2 learners with an average of 90.83 words per text ($SD= 22.06$) as compared to their HL counterparts who averaged 100.75 words per text ($SD= 26.43$), $t(194.6)= -3.57, p = 0.0004; d = -0.41$.

With respect to the total number of errors revised by each group, the WCF group, as a result of the feedback treatment, received a total of 955 corrections and averaged 4.96 self-revisions per text during the treatment period. The No-WCF group, on the other hand, completed a total of 646 revisions without corrective feedback, with an average of 3.34 self-revisions per
text. As shown in Figure 4.1 below, the total number of errors revised during the revision sessions was considerably higher in the WCF group than in the No-WCF in 11 out of the 12 grammatical categories, including gender and non-canonical gender marking, spelling, word choice, verb conjugation, number agreement, omission of definite articles, relative pronoun ‘que’, accusative marker ‘a’, and tense; whereas the no-WCF group averaged more self-revisions related to accent marking than the WCF group.

![Figure 4.1](image)

**Figure 4.1**

*Total Number of Revisions through all sessions (WCF and No-WCF Groups)*

**Notes.** WCF = revision with corrective feedback; No WCF = revision without corrective feedback; Total Number = number of corrections after all revision sessions

A series of independent samples t-tests were computed to further compare the total number of revisions made by each group for each grammatical category. The results revealed that the WCF group revised significantly more errors than the No-WCF group in canonical gender marking, \( t(34)=3.02, p=0.005 \), with equal variances assumed, and also in non-canonical gender marking, \( t(19.5)=2.81, p=0.011 \); omission of definite article, \( t(17.9)=2.87, p=0.07 \);
omission of accusative marker ‘a’, \( t(24)= 2.38, p = 0.026 \), and subjunctive mood, \( t(20)= 3.04, p = 0.007 \), (equal variances not assumed). However, a comparison between the two groups revealed no significant difference in the total number of revisions in the seven other categories - spelling, accent marking, word choice, verb conjugation, number agreement, omission of ‘que’ and tense (equal variances not assumed); \( p > 0.05 \).

4.2. Research Question 1: Comparison between revision with and without comprehensive WCF in terms accuracy development and retention of L2 forms

Research question 1 asked: Do students who do revision with and without comprehensive WCF differ in terms of accuracy development and retention of specific linguistic forms over time? Specifically, are there any differences between the WCF and No-WCF conditions in terms of error revision and accurate production over time of (a) canonical gender marking, (b) non-canonical/irregular gender marking, (c) definite articles in obligatory contexts, and (d) the present subjunctive mood?

As indicated in Chapter 3, these four linguistic features were chosen for various reasons. First, as shown in Figure 4.1, the occurrence and frequency of error production and revision were relatively high in all four categories, which enabled their production to be tracked over time and be compared between the two groups. In addition, errors in these four categories are recurrent in the production of L2 and HL learners of Spanish whose L1 is English, regardless of their proficiency level, and, as SLA research has shown, these forms also vary in how they are acquired (i.e., rule-based or lexically stored) and also in their degree of complexity or difficulty (i.e., canonical gender marking vs. present subjunctive). As a result, the individual analyses of these four distinctive linguistic forms, in contrast to previous studies that have combined
multiple linguistic features into broad categories, as in Van Beuningan et al. (2012) who provided a ratio for ‘grammatical’ errors (i.e., number of article errors, inflectional errors, word order errors, omissions of necessary elements, additions of non necessary elements, pronominal errors, and other grammatical errors) will provide a clear understanding regarding the extent to which both revision with and without comprehensive WCF can treat specific error types and facilitate L2 acquisition of such forms.

In the next sections, the group results for each of the four linguistic features will be compared in terms of (1) production, (2) number of errors (3) error revision and (4) accuracy development or retention in order to assess the effects of the intervention.

4.2.1. Research question 1a: Canonical Gender Marking

The following sub-sections compare the results of the WCF and No-WCF groups in terms of total production, error revision and accuracy development of canonical gender marking. All instances of canonical gender-marking were classified as either cases of gender assignment or gender agreement.

**Canonical Gender Marking Production (WCF and No-WCF)**

The results indicate that the production and distribution of canonical gender marking was comparable between the WCF and No-WCF groups in both the production of canonical gender assignment and canonical gender agreement, as shown Figure 4.2.
Figure 4.2
Distribution of Canonical Gender Marking Production (WCF and No-WCF Groups)

<table>
<thead>
<tr>
<th></th>
<th>Assignment</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCF</td>
<td>82.7%</td>
<td>17.3%</td>
</tr>
<tr>
<td>N = 1491</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>84.4%</td>
<td>15.6%</td>
</tr>
<tr>
<td>N = 1434</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-WCF</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>17.3%</td>
</tr>
<tr>
<td>N = 266</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; N = number of gender marking cases; % = percentage of total gender marking

The participants in the WCF group produced a total of 1491 cases of canonical gender assignment, with an average of 7.97 instances per text (SD = 4.3), while the participants in the No-WCF group totaled 1434 cases of canonical gender assignment and averaged 7.51 cases per text (SD = 3.7). Additionally, the participants in the WCF group produced a total of 311 cases of canonical gender agreement, with an average of 1.65 instances per text (SD = 1.5), whereas the No-WCF group produced 266 cases of canonical gender agreement and averaged 1.39 cases (SD = 1.5) per text.

As shown in Figure 4.2, the frequency of cases of canonical gender assignment was considerably higher (82.7%) compared to canonical gender agreement (17.3%) in the WCF group. Similarly, the canonical gender marking distribution in the No-WCF group was also higher for gender assignment (84.4%) as compared to gender agreement (15.6%). These differences were expected since gender assignment (determiner + noun) is required every time NPs are produced in Spanish as compared to gender agreement with attributive and predicative adjectives, which typically appear less frequently in production.
Canonical Gender Marking Errors (WCF and No-WCF)

The results, shown in Figure 4.3, revealed that the total number and distribution of canonical gender marking errors between the WCF and No-WCF groups was roughly the same in terms of gender assignment and gender agreement.

Figure 4.3
Distribution of Canonical Gender Marking Errors (WCF and No-WCF Groups)

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; N = number of gender marking errors per type; % = percentage of total gender marking per type

The participants in the WCF group produced a total of 61 canonical gender assignment errors, which ranged from 0 to 4 errors per text, with an average of 0.32 errors, (SD= 0.63) whereas the participants in the No-WCF group totaled 50 errors of canonical gender assignment, with a range from 0 to 5 errors per text, and an average of 0.26 errors (SD= 0.69).

The participants in the WCF group totaled 43 errors of canonical gender agreement, ranging from 0 to 5 errors per text with an average 0.23 errors (SD= 0.543). On the other hand, the No-WCF group produced 33 canonical gender agreement errors, ranging from 0 to 3 errors per text, and an average of 0.17 errors (SD= 0.44) per text.

A series of independent samples t-tests were further computed to assess whether the differences in the total number of canonical gender marking errors between the WCF and No-
WCF groups were statistically significant. The t-test results revealed that there were no significant differences between the WCF and No-WCF conditions with respect to the total number of canonical gender assignment errors, $t(376)= 0.91, p = 0.36$; or canonical gender agreement errors, $t(360)= 1.08, p = 0.28$.

Canonical Gender Marking Revision (WCF and No-WCF)

As part of the treatment, the WCF group received a total of 77 corrections for their canonical gender marking errors; while the No-WCF group, the control condition, did not receive any grammar feedback on their gender marking errors, as shown in Table 4.1.

<table>
<thead>
<tr>
<th></th>
<th>WCF</th>
<th>No WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment</td>
<td>N=43</td>
<td>N=0</td>
</tr>
<tr>
<td>Agreement</td>
<td>N=33</td>
<td>N=0</td>
</tr>
<tr>
<td>(M=0.28; SD=0.6)</td>
<td>(M=0.22; SD= 0.51)</td>
<td></td>
</tr>
</tbody>
</table>

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback: N = number of gender marking corrections per type; M= Mean number of corrections per text; SD = Standard Deviation

The results show that the WCF group received grammar feedback on a total of 43 gender assignment errors, ranging from 0 to 4 corrections per text with a mean of 0.28 corrections per text (SD=0.6), during the treatment period; while the No-WCF group received no error corrections on their canonical gender assignment errors. The WCF group also received grammar feedback on a total of 33 canonical gender agreement errors, ranging from 0 to 5 corrections per text with an average of 0.22 corrections per text (SD= 0.51); while the participants in the No-
WCF group, as shown above in Table 4.1, received no grammar corrections on their canonical gender agreement errors.

With respect to the total number and proportion of canonical gender marking errors revised by each group, the results revealed that the WCF group was able to revise more errors than the No-WCF group, as shown in Figure 4.4.

Figure 4.4
Total number of Revisions and Percentage of Total Revisions with respect to Total Errors (WCF and No-WCF Groups)

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; N = number of correct gender marking revisions per type; % = percentage of total correct revisions with respect to total errors

The participants in the WCF group had a total of 42 correct canonical gender assignment revisions, ranging from 0 to 4 revisions per text, and a mean of 0.27 revisions (SD=0.6), which involved 80% of all of the canonical gender assignment errors produced during the treatment period. In terms of canonical gender agreement, the WCF group totaled 32 revisions of canonical gender agreement errors, ranging from 0 to 2 revisions per text (M=0.21; SD= 0.5), which corresponded to the revision of 80.8% of all of their canonical gender agreement errors.

On the other hand, the No-WCF group provided fewer canonical gender marking revisions, including a total of 11 revisions of canonical gender assignment errors, ranging from 0
to 3 revisions per text (M= 0.07; SD= 0.34), entailing 28.2% of all of their canonical gender assignment errors. In addition, the No-WCF group revised a total of 11 canonical gender agreement errors (M=0.07; SD=0.26), which represented the revision of 25.8% of their gender agreement errors.

In order to test the efficacy of the WCF treatment on canonical gender marking error revision, an independent samples t-test was conducted to compare the total number of revisions between the two group conditions for each type. The t-test results indicated that the participants in the WCF group were able to correctly revise significantly more canonical gender marking errors than the No-WCF group, gender assignment, $t(239.53)= 3.69$, $p = 0.000$ and canonical gender agreement, $t(211.38)= 3.41$, $p = 0.001$. Furthermore, the results revealed that the WCF treatment had a medium-sized effect ($d = 0.43$) for the revision of canonical gender assignment errors and a low-medium sized effect ($d = 0.36$) for the revision of canonical gender agreement errors.

**Canonical Gender Marking Accuracy Development (WCF and No-WCF)**

To determine the effects of revision with and without WCF on the accuracy development of canonical gender marking, the percentage of correct use of gender assignment and gender agreement in each session was calculated and individual accuracy linear trends for each type of canonical gender marking were computed for each group condition.

**Accuracy Development of Canonical Gender Assignment (WCF and No-WCF)**

First, the initial canonical gender assignment accuracy results between the two group conditions were compared in order to observe whether any differences existed before the treatment. The results showed comparable results between the WCF group ($M= 94.1$, $SD=11.1$) and the No-WCF group ($M= 95.5$, $SD=12.6$) in their initial accuracy levels before the treatment.
Descriptive results on the accurate production of canonical gender assignment for each of the 11 sessions are shown in Table 4.2.

Table 4.2

*Canonical gender assignment* accuracy results per session

<table>
<thead>
<tr>
<th>Session</th>
<th>WCF Mean</th>
<th>WCF SD</th>
<th>No WCF Mean</th>
<th>No WCF SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>94.1</td>
<td>11.1</td>
<td>1</td>
<td>95.5</td>
</tr>
<tr>
<td>2</td>
<td>98.5</td>
<td>4.1</td>
<td>2</td>
<td>96.7</td>
</tr>
<tr>
<td>3</td>
<td>93.8</td>
<td>12.3</td>
<td>3</td>
<td>100.0</td>
</tr>
<tr>
<td>4</td>
<td>92.8</td>
<td>15.0</td>
<td>4</td>
<td>98.3</td>
</tr>
<tr>
<td>5</td>
<td>96.6</td>
<td>6.1</td>
<td>5</td>
<td>95.1</td>
</tr>
<tr>
<td>6</td>
<td>95.9</td>
<td>6.1</td>
<td>6</td>
<td>95.2</td>
</tr>
<tr>
<td>7</td>
<td>93.7</td>
<td>10.2</td>
<td>7</td>
<td>96.6</td>
</tr>
<tr>
<td>8</td>
<td>95.6</td>
<td>13.3</td>
<td>8</td>
<td>89.3</td>
</tr>
<tr>
<td>9</td>
<td>93.4</td>
<td>12.2</td>
<td>9</td>
<td>96.8</td>
</tr>
<tr>
<td>10</td>
<td>94.7</td>
<td>9.9</td>
<td>10</td>
<td>97.4</td>
</tr>
<tr>
<td>11</td>
<td>97.5</td>
<td>7.1</td>
<td>11</td>
<td>91.3</td>
</tr>
<tr>
<td>Total</td>
<td>95.1</td>
<td>10.1</td>
<td>Total</td>
<td>95.7</td>
</tr>
</tbody>
</table>

*Notes.* Session = class session; Mean = percentage of accuracy; N = number of cases; WCF = revision plus corrective feedback group; No WCF = revision without corrective feedback group; SD = standard deviation.

The results in Table 4.2 show variation in the mean production of canonical gender assignment in both groups throughout all of the sessions, which ranged from 92.8% to 98.5% of accuracy in the WCF (M=95.12, SD=10.1); and from 91.3% to 100% in the No-WCF (M=95.7; SD= 13.1).

In addition, a scatter plot diagram, shown in Figure 4.5 below, was generated in order to compare the effects of revision with and without WCF on the accuracy development of canonical gender assignment during the data collection period.
The accuracy trends indicate that the two groups followed different trajectories throughout the 11 sessions. The accurate production of canonical gender assignment in the WCF group remained stable over time, while it slightly decreased in the No-WCF group.

At the individual level, the percentage of students in each group that increased, decreased, and remained stable in terms of accuracy in canonical gender assignment was calculated and is shown in Table 4.3 below.

Table 4.3
Direction of individual developmental trends Canonical Gender Assignment (WCF and No-WCF groups)

<table>
<thead>
<tr>
<th>Trend</th>
<th>WCF</th>
<th>No-WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase</td>
<td>6</td>
<td>33.3%</td>
</tr>
</tbody>
</table>

---

2 Scatter plot diagrams with individual accuracy trends for each group are included in Appendix E
Table 4.3 (cont.)

<table>
<thead>
<tr>
<th></th>
<th>Decrease</th>
<th>Stable</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>7</td>
<td>5</td>
<td>18</td>
</tr>
<tr>
<td>%</td>
<td>38.9%</td>
<td>27.8%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Notes. WCF = revision plus corrective feedback group; No WCF = revision without corrective feedback group; % = percentage of total participants

The individual results showed that a higher percentage of participants in the WCF group (N=6; 33.3%) increased their accurate production of canonical gender assignment over time compared to the participants in the No-WCF group (N=3; 16.7%), but there were, however, more participants in the WCF group whose canonical gender assignment accuracy also decreased over time (N=7; 38.9%) than in the No-WCF group (N=3; 16.7%). Lastly, the individual results revealed that a majority of participants in the non-WCF group showed stable trends over time (N=10; 55.5%) as compared to the WCF group (N=5; 27.8%).

Accuracy Development of Canonical Gender Agreement (WCF and No-WCF)

A comparison of the initial accuracy results on canonical gender agreement between the WCF ($M= 95$, $SD=14.01$) and No-WCF groups ($M= 86.37$, $SD=23.35$) revealed no significant difference between the two groups before the beginning of the treatment, $t(15.498)= 1.083$, $p = 0.295$. Descriptive statistics on the production of canonical gender agreement for each of the eleven sessions are included in Table 4.4 below.
Table 4.4
Canonical gender agreement accuracy results per session

<table>
<thead>
<tr>
<th>Session</th>
<th>WCF Mean</th>
<th>SD</th>
<th>No WCF Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95.0</td>
<td>14.0</td>
<td>1</td>
<td>86.4</td>
</tr>
<tr>
<td>2</td>
<td>83.3</td>
<td>29.1</td>
<td>2</td>
<td>85.1</td>
</tr>
<tr>
<td>3</td>
<td>90.7</td>
<td>16.4</td>
<td>3</td>
<td>84.5</td>
</tr>
<tr>
<td>4</td>
<td>82.2</td>
<td>37.5</td>
<td>4</td>
<td>93.6</td>
</tr>
<tr>
<td>5</td>
<td>98.2</td>
<td>6.0</td>
<td>5</td>
<td>96.4</td>
</tr>
<tr>
<td>6</td>
<td>78.2</td>
<td>38.1</td>
<td>6</td>
<td>84.7</td>
</tr>
<tr>
<td>7</td>
<td>80.6</td>
<td>36.1</td>
<td>7</td>
<td>73.6</td>
</tr>
<tr>
<td>8</td>
<td>58.3</td>
<td>50.0</td>
<td>8</td>
<td>85.7</td>
</tr>
<tr>
<td>9</td>
<td>86.3</td>
<td>18.0</td>
<td>9</td>
<td>86.5</td>
</tr>
<tr>
<td>10</td>
<td>96.4</td>
<td>13.4</td>
<td>10</td>
<td>90.0</td>
</tr>
<tr>
<td>11</td>
<td>85.0</td>
<td>33.7</td>
<td>11</td>
<td>90.0</td>
</tr>
<tr>
<td>Total</td>
<td>86.8</td>
<td>27.7</td>
<td>Total</td>
<td>87.1</td>
</tr>
</tbody>
</table>

Notes. Session = class session; Mean = group percentage of accuracy
WCF = revision plus corrective feedback group; No WCF = revision without corrective feedback group; SD = standard deviation.

The results show great variation in the accurate group production of canonical gender agreement per session, ranging from 58.3% to 98.2% accuracy in the WCF group, and 73.6% to 96.4% in the No-WCF group. The results in Table 4.4 also revealed comparable results in the total average of gender agreement accuracy means throughout all sessions between the WCF group (M=86.8; SD=27.7) and the No-WCF group (M=87.1; SD=27.1).

The scatter plot diagram displayed in Figure 4.6 below indicates that the two groups followed different trajectories in terms of canonical gender agreement accuracy development.
As shown in the diagram, the accurate production of canonical gender agreement in the WCF group slightly decreased over time, whereas the group trend in the no-feedback condition remained stable over time.

At the individual level, the results revealed great within group variation in terms of accuracy development in the two groups, as shown in Table 4.5 below.

Table 4.5
Direction of individual developmental trends Canonical Gender Agreement (WCF and No-WCF groups)

<table>
<thead>
<tr>
<th>Trend</th>
<th>WCF</th>
<th>No-WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Decrease</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Stable</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>18</td>
</tr>
</tbody>
</table>

Notes. WCF = revision plus corrective feedback group; No WCF = revision without corrective feedback group; % = percentage of total participants
The individual accuracy patterns revealed that there were fewer participants in the WCF group whose canonical gender agreement accuracy increased over time (N=6; 33.3%) as compared to the participants in the No-WCF group (N=8; 44.4%). Additionally, a larger number of participants in the WCF group experienced a decrease in canonical gender agreement accuracy over time (N=7; 38.9%) in comparison to the participants in the No-WCF group (N=5; 27.8%). Lastly, the results indicate that the same number of participants in each group (N=5; 27.8%) showed stable trends in terms of gender agreement development during the data collection period (N=5; 27.8%).

4.2.2. Research question 1b: Non-Canonical Gender Assignment

The following section compares the total production, error revision and accuracy development of non-canonical gender marking between the two group conditions. For the purpose of the analysis, all instances of non-canonical gender-marking were classified as either cases of gender assignment or gender agreement.

Non-Canonical Gender Marking Production (WCF and No-WCF)

A comparison in terms of non-canonical gender marking revealed that the total number and distribution of gender assignment and agreement cases were similar in the WCF and No-WCF groups, as shown Figure 4.7.
The participants in the WCF group produced a total of 1102 cases of non-canonical gender assignment, with an average of 5.89 cases per text (SD= 3.6), and a total of 120 instances of non-canonical gender agreement, with an average of 0.65 cases per text (SD=1.0). More cases of canonical gender marking were found in the production of the No-WCF group, including a total of 1170 cases of non-canonical gender assignment, with an average of 6.16 instances per text (SD=3.4) and 160 cases of non-canonical gender agreement, with an average of 0.85 cases per text (SD= 1.1). In terms of gender marking distribution, the frequency of production of non-canonical gender assignment was higher in both the WCF (90.2%) and No-WCF (88%) groups, as compared to the production of non-canonical gender agreement, WCF group (9.8.%) and No-WCF (12%).

An independent samples t-test was computed in order to assess whether the differences in the total average production of non-canonical gender marking were significant between the two groups. The t-test results found no significant differences between the two conditions in terms of
the mean production of gender assignment, \( t(375) = -0.73, p = 0.47 \); or gender agreement, \( t(372) = -1.82, p = 0.069 \).

Non-Canonical Gender Marking Errors (WCF and No-WCF)

The results revealed some differences between the WCF and No-WCF groups in the total number and distribution of canonical gender marking errors, as shown in Figure 4.8 below.

Figure 4.8
Distribution of Non-Canonical Gender Marking Errors per type (WCF and No-WCF Groups)

<table>
<thead>
<tr>
<th></th>
<th>Assignment</th>
<th>Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCF</td>
<td>69.9%</td>
<td>30.1%</td>
</tr>
<tr>
<td>No-WCF</td>
<td>78.0%</td>
<td>22.0%</td>
</tr>
</tbody>
</table>

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; \( N \) = number of gender marking errors per type; \( \% \) = percentage of total gender marking per type.

The participants in the WCF group produced a total of 93 non-canonical gender assignment errors, ranging from 0 to 6 errors per text, with an average of 0.50 errors (SD = 1.007), whereas the participants in the No-WCF group totaled 103 errors of canonical gender assignment, with a similar range of errors from 0 to 6, and an average of 0.54 errors per text (SD = 1.016). With respect to non-canonical gender agreement, the WCF group totaled 40 errors of canonical gender agreement, ranging from 0 to 6 errors per text, and averaged 0.22 per text (SD = 0.65). On the other hand, the No-WCF group produced fewer non-canonical gender agreement errors (N=29, ranging from 0 to 2 errors per text with an average of 0.15 errors (SD = 0.40).
A series of independent samples t-tests were further calculated to assess whether, in terms of total gender marking errors, differences between the WCF and No-WCF groups were statistically significant. The t-test results revealed no significant differences between the two conditions in the total number of non-canonical gender assignment errors, $t(375) = -0.43, p = 0.67$; or non-canonical gender agreement errors, $t(308.3) = 1.12, p = 0.26$.

**Non-Canonical Gender Marking revision (WCF and No-WCF)**

As a result of the treatment, the participants in the WCF group received grammar correction on a total of 94 non-canonical gender marking errors, including a total of 64 corrections of gender assignment errors and 30 corrections of gender agreement errors. The participants in the No-WCF, the control group, did not receive any corrections on their non-canonical gender assignment (N=0) and non-canonical gender agreement (N=0) errors, as shown in Table 4.6 below.

**Table 4.6**

*Total number of feedback corrections received on Non-Canonical Gender Marking errors (WCF and No-WCF groups)*

<table>
<thead>
<tr>
<th>WCF Group</th>
<th>No WCF Group</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assignment</strong></td>
<td><strong>Agreement</strong></td>
</tr>
<tr>
<td>N=64</td>
<td>N=30</td>
</tr>
<tr>
<td>(M=1.42; SD=1.36)</td>
<td>(M=0.67; SD= 0.71)</td>
</tr>
</tbody>
</table>

*Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; N = number of gender marking corrections per type; M= Mean number of corrections per text; SD = Standard Deviation*
With respect to the number of errors revised by each group, the WCF group totaled considerably more revisions of their non-canonical gender marking errors ($N=72$) than the No-WCF group condition ($N=22$), as shown in Figure 4.9.

**Figure 4.9.**
*Total number of Revisions and Percentage of Total Revisions with respect to Total Errors (WCF and No-WCF Groups)*

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; N = number of correct gender marking revisions per type; % = percentage of total correct revisions with respect to total errors

The results revealed that the participants in the WCF group had a total of 51 correct revisions, ranging from 0 to 6 revisions per text ($M=1.13; SD=1.26$) on the feedback they received on non-canonical gender assignment, which involved 64.6% of all of the non-canonical gender assignment errors produced during the treatment period ($N=79$); in addition to 21 revisions of non-canonical gender agreement errors, ranging from 0 to 2 revisions per text ($M=0.47; SD=0.7$), which represented the revision of 66.3% of their non-canonical gender agreement errors during the treatment period ($N=33$). On the other hand, the No-WCF group provided a total of 17 revisions of non-canonical gender assignment errors, ranging from 0 to 3 revisions per text ($M=0.29; SD=0.6$), which entailed 18.9% of all of their assignment errors ($N=90$), and they also revised a total of 5 non-canonical gender agreement errors, with a range of
0 to 1 revisions per text (M=0.09; 0.28), which corresponded to 29.4% of all gender agreement errors produced during the treatment period (N=19). A mean comparison between the two groups revealed that the WCF group provided significantly more revisions of non-canonical gender assignment errors, \( t(59.532)= 4.15, p = 0.0003 \); and non-canonical gender agreement errors, \( t(55.37)= 3.46, p = 0.001 \), than their No-WCF counterparts.

**Accurate Production of Non-Canonical Gender Marking Forms after Revision**

The effects of revision with and without WCF on the accurate production of non-canonical gender marking was measured by tracking the production of specific forms that were previously revised (or not revised) and subsequently used by the same participant in new pieces of writing. The total number of non-canonical gender marking forms that were traceable, that is, produced by the same participant in two or more different texts, is included in Table 4.7 below.

<table>
<thead>
<tr>
<th>WCF Group (n= 8)</th>
<th>No WCF Group (n= 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment Cases</td>
<td>Assignment Cases</td>
</tr>
<tr>
<td>N= 20</td>
<td>N= 23</td>
</tr>
</tbody>
</table>

*Notes.* WCF = revision with corrective feedback; No WCF = revision without corrective feedback; Subjects = participants per group; Cases: individual cases of non-canonical gender assignment; Mean = average number of cases per participant; SD = standard deviation.

The results in Table 4.7 show that a total of 43 non-canonical gender marking errors were traceable over time, all of which consisted of gender assignment forms, including 20 non-canonical gender assignment forms found in the written production of participants in the WCF group (N=8), and 23 forms in the production of participants in the No-WCF group (N=6).
All of the traceable non-canonical gender assignment forms were further classified according to whether the same revised form was (a) immediately produced by the same learner in the following text, (b) produced by the same learner in a later but not immediate text, or (c) produced in two (or more) different texts by the same learner. The descriptive statistics in Table 4.8 show the results of revision and no revision on the correct production of the same form in subsequent production.

Table 4.8
*Effects of revision with and without WCF on the accurate production of non-canonical gender assignment forms*

<table>
<thead>
<tr>
<th>WCF Group (n= 8)</th>
<th>No WCF Group (n= 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revision + Immediate Production</td>
<td>83.3% (N=6)</td>
</tr>
<tr>
<td>Revision + Later Production</td>
<td>57.1% (N=7)</td>
</tr>
<tr>
<td>Revision + Production in two different texts</td>
<td>14.3% (N=7)</td>
</tr>
<tr>
<td>No revision + Immediate Production</td>
<td>- (N= 0)</td>
</tr>
<tr>
<td>No revision + Later Production</td>
<td>- (N= 0)</td>
</tr>
<tr>
<td>No Revision + Immediate + Production in two different texts</td>
<td>- (N= 0)</td>
</tr>
</tbody>
</table>

*Notes.* % = Percentage of correct use of non-canonical gender assignment forms; N = number of non-canonical gender assignment forms; Revision = a non-canonical gender assignment form was corrected during the revision session; No Revision = the correct non-canonical gender assignment was not provided during the revision session; Immediate Production = the same non-canonical gender assignment form was used after the revision stage; Production in two different texts = the same non-canonical gender assignment form was used in two different texts; Later Production = the same non-canonical gender assignment form was used in a subsequent, but not immediate, text.
The accuracy results in the WCF group indicate that the highest rate of correct use corresponded to those forms that were revised and subsequently used in immediate production (83.3%). Conversely, the accurate production of revised forms that were used in later, but not immediate production was lower (57.1%) and the correct use of revised forms further decreased when the same form was produced in a second texts (14.3%), as shown in Figure 4.10 below.

With respect to the No-WCF group, the results revealed that the highest rate of accuracy corresponded to those forms that were revised and used in later production (75%), while the accurate production of a revised form in a second text was the (0%) although the sample only included one case (N=1). Furthermore, only in the No-WCF group there were instances of errors that were not revised but used correctly in new pieces of writing, including the correct use of non-canonical forms in immediate production (60%), and later production (23%). Finally, no cases of non-revised forms used in two different texts were found in the data.
4.2.3. Research question 1c: Definite Articles in Obligatory Contexts

The following sections present a comparison between the WCF and No-WCF groups in terms of total production, error revision and accuracy development of definite articles in obligatory contexts.

Production of Definite Articles in Obligatory Contexts (WCF and No-WCF)

Descriptive statistics on the total number of cases of definite articles in obligatory contexts for each condition are included in Figure 4.11. The results show that the total occurrence of definite articles in obligatory contexts was comparable between the WCF group and the No-WCF group.

Figure 4.11
Total occurrence of definite articles in obligatory contexts (WCF and No-WCF)

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; M = mean number of definite articles in obligatory contexts per text.

A total of 892 cases of definite articles in obligatory contexts were found in the production of the WCF group, which averaged 4.51 instances per text (SD= 3.2). Additionally, a total of 881 cases that required the use of definite articles were found in the written production of
the No-WCF group, with an average of 4.45 instances per text (SD= 2.7). A t-test comparison of the occurrence of definite articles in obligatory contexts between the two groups revealed no significant differences, $t(382.5)= 1.87, p = 0.85$.

**Definite Article Omission Errors (WCF and No-WCF)**

In terms of error production, the results indicate that the participants in the WCF group made fewer errors of omission of definite articles than their counterparts in the No-WCF group, as shown in Figure 4.12 below.

Figure 4.12  
*Total errors of definite article omission in obligatory contexts (WCF and No-WCF)*

The participants in the WCF group produced a total of 96 errors of definite article omission, ranging from 0 to 6 errors per text, with an average of 0.48 errors (SD=1.1). The No-WCF group made a total of 187 omission errors, which ranged from 0 to 9 errors per text, and
averaged 0.94 errors (SD=1.6). A t-test comparison revealed that this difference was significant, with the WCF group producing significantly fewer errors than the No-WCF group, \( t(351.5) = -3.27, p = 0.001, d=0.34 \).

**Revision of Definite Article Omission errors (WCF and No-WCF)**

The descriptive results displayed in Table 4.9 show that, as part of the treatment, the WCF group received a total of 82 corrections (M=0.55; SD=1.1) for their article omission errors, which ranged from 0 to 6 corrections per text. The participants in the No-WCF group, the control condition, did not receive any corrections on their omission errors (N=0).

<table>
<thead>
<tr>
<th>WCF Group</th>
<th>No WCF Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>N= 82</td>
<td>N=0</td>
</tr>
<tr>
<td>(M= 0.55; SD= 1.1)</td>
<td></td>
</tr>
</tbody>
</table>

*Notes.* WCF = revision with corrective feedback; No WCF = revision without corrective feedback; N = number of gender marking corrections per type; M= Mean number of corrections per text; SD = Standard Deviation

In terms of the total number and proportion of article omission errors successfully revised by each group, the results indicate that the WCF group provided a larger number of revisions than the No-WCF group, as shown in Figure 4.13 below.
The WCF group revised a total of 73 errors, ranging from 0 to 6 revisions per text, with an average of 0.49 cases per text (SD= 1.0), which comprised 94.85% of all the errors produced during the treatment period. The participants in the No-WCF revised a total of 8 errors, ranging from 0 to 2 revisions per text, with an average of 0.12 cases (SD=0.37) that involved the revision of 6.1% of all their article omission errors. Since the error t-test comparison revealed that the WCF and No-WCF groups significantly differed in the total number of article omission errors, a chi-square test was calculated to determine whether there was a significant difference between the two groups in terms of percentage of errors revised with respect to the total number of errors. The chi-square results revealed that the proportion of errors correctly revised by the WCF group was significantly higher as compared to the No-WCF group, $\chi^2 = 78.43, p = 0.00, d= 0.78$. 

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; N = number of correct revisions of article omission; % = percentage of total correct revisions with respect to total errors
**Accuracy Development of Definite Articles in Obligatory Contexts (WCF and No-WCF)**

The results revealed variation in the accurate production of definite articles in obligatory context throughout all of the sessions in the WCF and No-WCF groups, as shown in Table 4.10.

Table 4.10

*Accurate use of Definite Articles in Obligatory Contexts per session*

<table>
<thead>
<tr>
<th>Session</th>
<th>Subjects</th>
<th>N</th>
<th>Percent</th>
<th>SD</th>
<th>Session</th>
<th>Subjects</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No WCF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>18</td>
<td>69.2</td>
<td>48.0</td>
<td>1</td>
<td>6</td>
<td>19</td>
<td>83.3</td>
<td>40.8</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>43</td>
<td>76.7</td>
<td>40.7</td>
<td>2</td>
<td>12</td>
<td>50</td>
<td>64.2</td>
<td>35.1</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>52</td>
<td>89.7</td>
<td>27.1</td>
<td>3</td>
<td>18</td>
<td>66</td>
<td>86.7</td>
<td>24.9</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>131</td>
<td>78.1</td>
<td>36.6</td>
<td>4</td>
<td>17</td>
<td>107</td>
<td>81.6</td>
<td>28.5</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>118</td>
<td>84.8</td>
<td>32.5</td>
<td>5</td>
<td>17</td>
<td>93</td>
<td>77.1</td>
<td>34.1</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>114</td>
<td>84.2</td>
<td>31.6</td>
<td>6</td>
<td>18</td>
<td>106</td>
<td>84.5</td>
<td>20.8</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
<td>90</td>
<td>95.0</td>
<td>13.0</td>
<td>7</td>
<td>18</td>
<td>96</td>
<td>89.8</td>
<td>15.1</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>44</td>
<td>81.8</td>
<td>36.3</td>
<td>8</td>
<td>18</td>
<td>58</td>
<td>80.0</td>
<td>34.4</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>71</td>
<td>83.3</td>
<td>24.3</td>
<td>9</td>
<td>18</td>
<td>78</td>
<td>75.1</td>
<td>30.4</td>
</tr>
<tr>
<td>10</td>
<td>18</td>
<td>133</td>
<td>93.9</td>
<td>10.6</td>
<td>10</td>
<td>16</td>
<td>108</td>
<td>74.8</td>
<td>31.9</td>
</tr>
<tr>
<td>11</td>
<td>17</td>
<td>78</td>
<td>97.7</td>
<td>6.8</td>
<td>11</td>
<td>18</td>
<td>101</td>
<td>82.2</td>
<td>29.8</td>
</tr>
<tr>
<td>Total</td>
<td>183</td>
<td>892</td>
<td>85.5</td>
<td>30.1</td>
<td>Total</td>
<td>176</td>
<td>882</td>
<td>80.3</td>
<td>29.1</td>
</tr>
</tbody>
</table>

Notes. Session = class session; Subjects = total participants per session; Mean = percentage of accuracy; N = number of cases; WCF = revision with corrective feedback group; No WCF = revision without corrective feedback group; SD = standard deviation.

The WCF group averaged 85.5% correct use of definite articles in obligatory contexts (SD=30.1), ranging from 69.2% to 97.7%. The No-WCF averaged 80.3% (SD=29.1) accuracy with scores that ranged from 64.2% to 89.8%.

A scatter plot diagram, displayed in Figure 4.14 below, was computed to compare the effects of the two conditions on the accurate production of definite articles in obligatory contexts over time.
The results indicate that the two groups followed different directions in the accurate production of definite articles in obligatory contexts, as the accuracy trend in the WCF group increased over time, while the developmental trajectory of the No-WCF group remained stable.

At the individual level, the percentage of participants in each group whose accuracy increased, decreased, and remained stable was calculated and is displayed in Table 4.11 below.

Table 4.11
Direction of individual developmental trends on the Use of Definite Articles in Obligatory Contexts (WCF and No-WCF groups)

<table>
<thead>
<tr>
<th>Trend</th>
<th>WCF</th>
<th>No-WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increase</td>
<td>8</td>
<td>44.4%</td>
</tr>
<tr>
<td>Decrease</td>
<td>4</td>
<td>22.2%</td>
</tr>
<tr>
<td>Stable</td>
<td>6</td>
<td>33.3%</td>
</tr>
</tbody>
</table>

Notes. WCF = revision plus corrective feedback group; No WCF = revision without corrective feedback group; N = number of participants; % = percentage of total participants
The individual accuracy trends revealed that the same number of participants in each group (N=8, 44%) increased their accurate use of definite articles in obligatory contexts. In addition, more participants in the No-WCF group experienced a decrease in accuracy over time (N=7; 38.9%) compared to the WCF group (N=4; 22.2%). The results also showed that there were more participants in the WCF group whose accuracy trends remained stable over time (N=6; 33.3%) as compared to the participants in the No-WCF group (N=3; 16.7%).

4.2.4. Research question 1d: Present of Subjunctive

The following section compares the WCF and No-WCF groups in terms of total production, error revision and accuracy development of the present subjunctive.

*Production of Present Subjunctive (WCF and No-WCF)*

All the cases that required the use of the present subjunctive were classified according to the type of subordinate clause, including nominal, adjectival and adverbial clauses. Given that nominal clauses were considerably more frequent in the written production of the participants in the two groups, all cases were further classified according to the modality and expression of the verb in the main clause, which included (a) possibility, (b) doubt, (c) opinion, (d) negation, (e) influence and volition). The results displayed in Figure 4.15 show similarities in the total number and distribution of structures that elicited the use of present subjunctive.
Figure 4.15
*Distribution of Present Subjunctive Occurrence per Type (WCF and No-WCF Groups)*

![Diagram showing distribution of present subjunctive occurrence per type]

**Notes.** WCF = revision with corrective feedback; No WCF = revision without corrective feedback; % = percentage of total subjunctive cases

A total of 81 structures classified into six different categories triggered the use of subjunctive in the production of the WCF group. The use of the present subjunctive in the WCF group was primarily elicited by verbs of denial and negation (N=30), which represented 37% of all of the cases, followed by impersonal structures expressing opinion, (N=20; 24.7%), verbs of influence/volition (N=8; 9.9%), adverbial conjunctions (N=8; 9.9%), unknown antecedent in adjectival clauses (N=8; 9.9%), and impersonal structures expressing possibility (N=7; 8.6%).

With respect to the No-WCF group, a total of 76 structures from seven different categories elicited the use of the present subjunctive, which was also largely triggered by verbs of denial and negation (N=32; 42.1%) and impersonal structures expressing opinion, (N=13; 17.1%), followed by unknown antecedent in adjectival clauses (N=11; 14.5%), adverbial
conjunctions (N=7; 9.2%) and verbs of influence and volition (N=6; 7.9%), impersonal structures expressing possibility (N=6; 7.9%), impersonal expressions with the pronoun se (N=2; 2.6%) and expressions of doubt (N=1; 1.3%).

*Present Subjunctive Errors (WCF and No-WCF)*

An independent samples t-test that compared the total number of the present subjunctive errors produced by the participants in the WCF and No-WCF groups across all the sessions showed no significant differences between the two conditions, \( t(189) = 0.574, p = 0.57 \).

A further comparison in terms of the categories that elicited the used of the present subjunctive revealed some similarities and differences in the total number of errors and accurate production of present subjunctive structures, shown in Table 4.12 below.

<table>
<thead>
<tr>
<th>Category</th>
<th>WCF</th>
<th>No-WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Influence/volition</td>
<td>6</td>
<td>25.0%</td>
</tr>
<tr>
<td>Negation</td>
<td>25</td>
<td>16.7%</td>
</tr>
<tr>
<td>Opinion</td>
<td>11</td>
<td>45.0%</td>
</tr>
<tr>
<td>Adverbal</td>
<td>2</td>
<td>75.0%</td>
</tr>
<tr>
<td>Adjectival</td>
<td>6</td>
<td>25.0%</td>
</tr>
<tr>
<td>Doubt</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Possibility</td>
<td>6</td>
<td>14.3%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>56</strong></td>
<td><strong>33.5%</strong></td>
</tr>
</tbody>
</table>

*Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; N = number of present subjunctive errors per type; % = percentage of accurate production.*

The WCF produced a total of 56 present subjunctive errors, using it correctly in 33.5% of cases. The participants in the No-WCF totaled 52 present subjunctive errors (37% accuracy).
The results indicate some variation in the accurate use of present subjunctive in nominal clauses between the two groups. The accurate production of the present subjunctive with verbs of influence/volition was slightly lower in the WCF group (25%) than in the No-WCF group (33.3% correctly produced). Similarly, the overall accurate production of the present subjunctive triggered by impersonal expressions of possibility was higher in the No-WCF group (50%) than in the WCF group (14.3%). Conversely, the production of the present subjunctive in nominal clauses elicited by verbs that express opinion was higher in the WCF group (45%) compared to the No-WCF (23%). The use of the present subjunctive in nominal clauses triggered by negation was slightly more accurate in the WCF group (16.7%) and the No-WCF group (12.2%).

In addition, the overall accurate production of the present subjunctive in adjectival clauses with an unknown antecedent was higher in the No-WCF group (54.3%) than in the WCF group (25%). The use of the present subjunctive in adverbial clauses was overall more accurate in the No-WCF group (85.7%) than in the WCF group (75%).

The results also revealed that there were more cases of errors of oversupply of the subjunctive, that is, the use of the present subjunctive where the use of the indicative is obligatory, in the WCF group than in the No-WCF group, as shown in Table 4.13.

<table>
<thead>
<tr>
<th>Category</th>
<th>WCF</th>
<th>No-WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oversupply of Subjunctive</td>
<td>N=20</td>
<td>N=11</td>
</tr>
</tbody>
</table>

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; N = number of cases of overgeneralized use of subjunctive
A comparison between the two groups showed that errors of oversuppliance of the present subjunctive forms were more frequent in the WCF group with a total of 20 forms that comprised 26.3% of all their errors related to the present subjunctive, as compared to the No-WCF group that totaled 11 errors of oversuppliance of present subjunctive, which in turn corresponded to 17.4% of all of their errors.

**Revision of Present Subjunctive errors (WCF and No-WCF)**

The results shown in Table 4.14 below provide an indication of the total and mean number of corrections that the participants in the WCF and No-WCF group received during the treatment period.

Table 4.14.

<table>
<thead>
<tr>
<th>Total number of feedback corrections received on Present Subjunctive errors (WCF and No-WCF groups)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCF Group</td>
</tr>
<tr>
<td>Corrections</td>
</tr>
<tr>
<td>N= 0.73 (SD = 0.44)</td>
</tr>
</tbody>
</table>

*Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback: N = number of gender marking corrections per type; M= Mean number of corrections per text; SD = Standard Deviation*

The WCF group received a total of 48 corrections of errors of the present subjunctive (M=0.74; SD= 0.44), ranging from 0 to 1 corrections per text, which represented 85.7% of all the errors produced during the treatment period, whereas the participants in the No-WCF group, that is the control condition, did not receive any grammar corrections for their present subjunctive errors (N=0).
With respect to the total number of revisions of present subjunctive errors, the proportion and total number of errors revised by the WCF group was higher than the No-WCF group, as shown in figure 4.16

**Figure 4.16**
*Total number of Present Subjunctive Error Revisions and Percentage of Total Revisions with respect to Total Errors (WCF and No-WCF Groups)*

![Bar chart showing total number of revisions and percentage for WCF and No-WCF groups.]

**Notes.** WCF = revision with corrective feedback; No WCF = revision without corrective feedback; N = number of correct revisions of present subjunctive errors; % = percentage of total correct revisions with respect to total errors

The participants in the WCF group revised 41 errors, or 73.2% of all of their subjunctive errors produced, with an average of 0.63 revisions (SD= 0.49) ranging from 0 to 1 revision per text. The No-WCF group revised a total of 5 present subjunctive errors, which entailed the revision of 9.6% of all of their errors produced during the treatment period, and averaged to 0.09 per text (SD= 0.28) with a range of 0 to 1 revision per text. A t-test comparison was computed to determine whether the difference in the total number of errors successfully revised between the
two groups was significant. The results showed that the total number of errors successfully revised was significantly higher in the WCF group compared to the No-WCF group, $t(106.2)=7.57, p=0.000, d=1.35$

**Accurate Production of Present Subjunctive Forms (WCF and No-WCF)**

As indicated in the methodology section, the effects of revision with and without WCF on the accurate production of specific present subjunctive forms was measured by tracking the production of erroneous forms that were revised (or not revised) and subsequently used by the same participant in new pieces of writing. The results in Table 4.15 below indicate that the written data of the No-WCF group (N=11) included a larger number of present subjunctive forms that were traceable over time as compared to WCF group (N=8).

<table>
<thead>
<tr>
<th></th>
<th>WCF Group (N=8)</th>
<th>No WCF Group (N=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learners</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cases</td>
<td>N= 13</td>
<td>N= 21</td>
</tr>
</tbody>
</table>

*Notes.* WCF = revision with corrective feedback; No WCF = revision without corrective feedback; Cases: individual cases of non-canonical gender assignment; Mean = average number of cases per participant; SD = standard deviation.

All of the traceable present subjunctive forms were classified according to whether the same revised form was (a) immediately produced by the same learner in the following text, (b) produced by the same learner in a later but not immediate text, or (c) produced by the same learner in two (or more) different texts. The results of the effect of revision with and without WCF on the accurate production of present subjunctive forms is shown in Table 4.16.
Table 4.16  
Effects of revision with and without WCF on the accurate production of present subjunctive forms

<table>
<thead>
<tr>
<th>WCF Group (n= 8)</th>
<th>No WCF Group (n= 11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revision + Immediate Production</td>
<td>- (N=0)</td>
</tr>
<tr>
<td>Revision + Later Production</td>
<td>0% (N=9)</td>
</tr>
<tr>
<td>Revision + Production in two different texts</td>
<td>- (N=0)</td>
</tr>
<tr>
<td>No revision + Immediate Production</td>
<td>0% (N= 2)</td>
</tr>
<tr>
<td>No revision + Later Production</td>
<td>0% (N= 2)</td>
</tr>
<tr>
<td>No Revision + Production in two different texts</td>
<td>- (N=0)</td>
</tr>
</tbody>
</table>

Notes. % = Percentage of correct use of present subjunctive; N = number of present subjunctive forms; Revision = a present subjunctive form was corrected during the revision session; No Revision = the present subjunctive form was not provided during the revision session; Immediate Production = the same present subjunctive form was used after the revision stage; Production in two different texts = the same present subjunctive form was used in two different texts; Later Production = the present subjunctive form was used in a subsequent, but not immediate, text.

Regarding the traceable forms in the WCF group that were produced in new texts, the results showed no instances of correct use (0%) of present subjunctive forms that were correctly revised and produced in a later text (N=9). With respect to the accurate production of unrevised present subjunctive forms, the results showed no correct use (0%) of the same forms in either immediate production (N=2) or later production (N=2), as in Figure 4.17 below.
Regarding the accurate production of revised forms in the No-WCF group, the results in Figure 4.17 also reveal no instances of correct use (0%) of present subjunctive forms that were revised and used in either immediate (N=1) or later texts (N=1). Additionally, no instances of correct use (0%) of present subjunctive forms were found in the No-WCF group of unrevised present subjunctive forms that were incorporated in either immediate (N=6) or later production (N=14).

4.3. Research question 2: Comparison between L2 and HL learners

A further analysis was conducted in order to answer research question #2, that is, to determine whether the L2 and HL learners in the WCF and No-WCF conditions differed in terms of production, error revision, and acquisition of canonical gender marking, non-canonical gender marking, use of definite articles in obligatory contexts and present subjunctive.
4.3.1. Research question 2a: Canonical Gender Marking

The following section compares the results of the L2 and HL learners within and across the two groups in terms of production, error revision and accuracy development of canonical gender assignment and canonical gender agreement.

Comparison of Canonical Gender Marking Production (L2 and HL)

The production results revealed that the total number of cases and frequency of each canonical gender marking type was comparable between the L2 and HL within and across the WCF and No WCF conditions, as shown in Figures 4.18 and 4.19 below.

![Graph showing percentage of canonical gender marking production per learner](image)

Notes. WCF = revision with corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of gender marking cases; % = percentage of total gender marking

The L2 learners in the WCF group produced a total of 1029 cases of canonical gender assignment, with an average of 8.23 instances per text (SD= 4.59), while their HL counterparts in the WCF group totaled 462 cases of canonical gender assignment and averaged 7.45 cases per text (SD=3.43). Additionally, the L2 learners in the WCF group totaled 193 cases of canonical gender agreement, with an average of 1.54 instances per text (SD=1.4), whereas the HL learners in the same condition produced a total of 118 cases of canonical gender agreement and averaged...
more cases of gender agreement per text (M=1.87; SD= 2.15). In terms of distribution, the frequency of canonical gender agreement production, as shown in Figure 4.18, was higher in the HL learners (21.6%) as compared to their L2 counterparts (15.8%) in the WCF group.

A set of independent samples t-tests was computed to compare whether the mean production differences between the L2 and HL learners in the WCF group were significant. The t-test results revealed no significant difference in the production of canonical gender assignment between the L2 learners and the HL learners in the WCF condition, \( t(185)= 1.18, p = 0.24 \), or in the production of canonical gender agreement, \( t(186)= -1.25, p = 0.21 \).

With respect to the participants in the No-WCF group, the results of the L2 and HL learners were comparable in terms of the production and distribution of canonical gender marking, as shown in Figure 4.19.

![Figure 4.19](image)

*Distribution of Canonical Gender Marking Production per learner (No WCF Group)*

**Notes.** No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of gender marking cases; % = percentage of total gender marking

The L2 learners in the No-WCF group produced a total of 939 cases of canonical gender assignment, with an average of 7.45 cases per text (SD= 3.8), while the HL learners in the No-WCF group totaled 495 cases of canonical gender assignment and averaged 7.73 instances per
text (SD=3.7). In addition, the L2 learners in the No-WCF condition produced a total of 171 instances of canonical gender agreement, with an average of 1.36 instances per text (SD=1.5), whereas their HL counterparts produced 95 cases of canonical gender agreement and averaged 1.48 cases per (SD=1.6). The distribution of canonical gender marking was also comparable between the L2 and HL learners in the No-WCF group, including canonical gender assignment (L2=84.6%; HL=83.9%) and gender agreement (L2=15.4%; HL=16.1%).

Independent samples t-tests were computed to determine whether the mean differences in gender marking production were significant among the L2 and HL learners in the No-WCF group. There were no significant differences between the L2 and HL learners, either in total production of gender assignment, \( t(188) = -0.50, p = 0.62 \), or gender agreement, \( t(188) = -0.55, p = 0.59 \).

**Comparison of Canonical Gender Marking Errors within group (L2 and HL)**

A comparison between the L2 and HL learners in the WCF and No-WCF groups revealed differences in the total number of canonical gender marking errors and distribution per type, as shown in Figure 4.20 and Figure 4.21.
The L2 learners in the WCF group produced a total of 52 canonical gender assignment errors, ranging from 0 to 4 errors per text, and also totaled 31 canonical gender agreement errors, which ranged from 0 to 2 errors per text. The results, as shown in Figure 4.20, revealed that assignment errors were more frequent (62.7%) than agreement errors (37.3%) in the production of the L2 learners. The HL learners in the WCF group produced a total of 9 gender assignment errors, with a range of 0 to 1 error per text, and 12 gender agreement errors, which ranged from 0 to 3 errors. Comparatively, there was a higher proportion of agreement errors (57.2%), as compared to assignment errors (42.8%) in the production of the HL learners.

Furthermore, a comparison between the two types of learners confirmed that the mean number of canonical gender assignment errors was significantly higher in the L2 learners (M=0.42; SD=0.71) than in the HL learners (M=0.14; SD=0.35) in the WCF group, t(185.9)=3.52, p=0.001, d=0.5. Conversely, no significant differences were found with respect to the number of canonical gender agreement errors between the L2 (M=0.25; SD=0.53) and HL learners (M=0.19; SD=0.56) in the WCF group, t(186)=0.68, p=0.495.

With respect to the learners in the No-WCF group, the results also revealed differences in the error production and distribution of canonical gender marking errors between the L2 and HL learners, as shown in Figure 4.21 below.
Figure 4.21
Distribution of Canonical Gender Marking Errors per learner type (WCF group)

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of gender marking errors; % = percentage of total gender marking errors per type.

The L2 learners in the No-WCF totaled 47 canonical gender assignment errors, ranging from 0 to 5 errors per text, and also produced 28 canonical gender agreement errors, with a range of 0 to 3 errors per text. Additionally, assignment errors were more recurrent (62.6%) in the production of the L2 learners than agreement errors (37.4%). The HL learners in the No-WCF group produced a total of 3 gender assignment errors, with a range of 0 to 1 error per text, and 5 gender agreement errors, which ranged from 0 to 1 error. Similar to the HL learners in the WCF group, the proportion of agreement errors (62.5%) was higher compared to assignment errors (37.5%) in the production of the HL learners in the No-WCF group.

A series of independent samples t-tests that compared the average number of gender marking errors revealed that the L2 learners in the No-WCF group produced significantly more gender assignment errors (M=0.37; SD=0.81) than their HL counterparts (M=0.05; SD=0.21), $t(155.8)= 4.25, p= 0.00003, d= 0.54$. Likewise, the average number of canonical gender agreement errors was also significantly higher in the L2 learners (M=0.22; SD=0.5) compared to HL learners (M=0.08; SD=0.27) in the No-WCF group, $t(187.4)= 2.56, p= 0.01, d= 0.35$. 
Comparison of Canonical Gender Marking Errors across groups (L2 and HL)

A series of independent samples t-tests were computed in order to determine whether the L2 and HL learners across the two groups differed in the average total number of canonical gender assignment and agreement errors. A comparison of the total number of canonical gender assignment errors produced by the L2 learners in the WCF group (M=0.44; SD=0.71) and the L2 learners in the No-WCF group (M=0.35; SD=0.78) showed no significant variation, \( t(202) = 0.84, p = 0.4 \). In terms of canonical gender agreement errors, there were no significant differences between the L2 learners in the WCF group (M=0.27; SD=0.57) and their L2 counterparts in the No-WCF group (M=0.25; SD=0.54), \( t(202) = 0.25, p = 0.075 \).

Similarly, the t-test results that compared the average total number of canonical gender marking errors between HL learners in the WCF group and the HL learners in the No-WCF group revealed no statistical differences in terms of the average number of either canonical gender assignment, \( t(87.03) = 1.38, p = 0.17 \); or canonical gender agreement, \( t(70.97) = 1.47, p = 0.14 \). Conversely, a comparison between the HL learners in the WCF condition and the L2 learners in the No-WCF showed significant differences in the error production of canonical gender marking. The t-test results revealed that the HL learners in the WCF group produced significantly fewer canonical gender assignment errors than the L2 learners in the No-WCF group, \( t(149.55) = -2.4, p = 0.19, d=0.17 \); however, no differences were found in terms of total error production of canonical gender agreement, \( t(151) = -2.0, p = 0.84 \).

Comparison of Revision of Canonical Gender Marking errors within group (L2 and HL)

A comparison between the L2 and HL learners in the WCF and No-WCF conditions showed differences in terms of the number of corrections and revisions of canonical gender marking errors. With respect to the number of corrections, the results, shown in Table 4.17
below, indicated that the L2 learners in the WCF group averaged considerably more corrections of canonical gender assignment errors (M=0.38; SD= 0.69) than their HL counterparts (M=0.08; SD= 0.27).

Table 4.17
Error Correction of Canonical Gender Assignment errors per group and learner (L2 and HL)

<table>
<thead>
<tr>
<th>Learner</th>
<th>WCF</th>
<th>No WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 (n=12)</td>
<td>N= 39</td>
<td>N= 0</td>
</tr>
<tr>
<td>HL (n=6)</td>
<td>(M=0.38; SD=0.69)</td>
<td>N= 0</td>
</tr>
<tr>
<td>L2 (n=12)</td>
<td>N= 4</td>
<td>N= 0</td>
</tr>
<tr>
<td>HL (n=6)</td>
<td>(M=0.08; SD= 0.27)</td>
<td></td>
</tr>
</tbody>
</table>

Notes. Errors = Total number of gender marking errors; Mean = average number of revisions per participant; WCF = revision with corrective feedback; No WCF = revision without corrective feedback: Revision*Errors = percentage of revisions with respect to total errors during treatment; SD = standard deviation

A t-test comparison was further computed to determine whether the mean differences in the number of gender assignment corrections received by the L2 and HL learners in the WCF group were significant. The results revealed that the differences between the two groups of learners were statistically significant, t(145.1)= 3.89, p= 0.000, and that the magnitude of difference was large, \(d= 0.71\), since L2 learners received many more error corrections than their HL counterparts.

Furthermore, the results, in Table 4.18 below, revealed that in the WCF group the L2 learners received on average more corrections of canonical gender agreement errors (M=0.24; SD=0.53) than the HL learners (M=0.18; SD=0.48); however a t-test comparison showed that the differences were not statistically significant \(t(151)= 0.67, p= 0.505\).
Table 4.18
*Error Correction of Canonical Gender Agreement* errors per group and learner (L2 and HL)

<table>
<thead>
<tr>
<th>Learner</th>
<th>WCF</th>
<th>No WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L2 (n=12)</td>
<td>HL (n=6)</td>
</tr>
<tr>
<td>Corrections</td>
<td>N= 24</td>
<td>N= 9</td>
</tr>
<tr>
<td></td>
<td>(M=0.24; SD=0.53)</td>
<td>(M=0.18; SD= 0.48)</td>
</tr>
<tr>
<td></td>
<td>L2 (n=12)</td>
<td>HL (n=6)</td>
</tr>
<tr>
<td></td>
<td>N= 0</td>
<td>N= 0</td>
</tr>
</tbody>
</table>

Notes. Errors = Total number of gender marking errors; Mean = average number of revisions per participant; WCF = revision with corrective feedback; No WCF = revision without corrective feedback; Revision*Errors = percentage of revisions with respect to total errors during treatment; SD = standard deviation

With respect to the revision of canonical gender marking errors, as shown in Figure 4.22 below, the L2 and HL learners in the WCF group were able to correctly revise a higher proportion of their total errors as compared to the L2 and HL learners in the No-WCF group.

![Figure 4.22](image)

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of correct gender marking revisions per type; % = percentage of total correct revisions with respect to total errors

The L2 learners in the WCF group provided the revision of a total of 38 gender assignment errors, which corresponded to 84% of all of their assignment errors produced during
the revision sessions, with a range of 0 to 4 revisions per text. The HL learners in the WCF group made a total of 4 revisions of canonical gender assignment errors with a range of 0 to 1 revisions per text, which comprised the revision of 57% of all of their errors. Given that the differences in the total number of canonical gender assignment errors between the HL and L2 learners was significant (\( p =< 0.05 \)), a chi-square test was performed to determine whether there was a significant difference between the percentage of errors revised with respect to the total number of errors made by each group. The results showed that the percentage of errors successfully revised by the L2 learners in the WCF group was significantly higher than their HL counterparts, \( \chi^2 = 5.17, p = 0.02, d = 0.88 \).

The canonical gender agreement results indicate that the L2 learners in the WCF group were able to revise a total of 23 errors, with a range of 0 to 4 errors, which consisted of 82% of all of their errors produced during the treatment. The HL learners revised a total of 9 gender agreement revisions, ranging from 0 to 2 revisions per text, which involved the revisions of 75% of all of their gender agreement errors. An independent samples t-test was computed to examine whether the L2 and HL learners significantly differed in the total number of revisions of canonical gender agreement. The t-test results found no significant differences between the L2 learners (\( M=0.23; SD=0.5 \)) and the HL learners (\( M=0.18; SD=0.4 \)) in the WCF group, \( t(151)=0.56, p = 0.576 \).

Regarding the participants in the No-WCF group, the proportion of canonical gender marking errors revised by the HL learners was higher as compared to the L2 learners, as shown in Figure 4.23 below.
Figure 4.23
*Total number and Percentage of Canonical Gender Marking Revisions with respect to Total Errors (L2 and HL Groups)*

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of correct gender marking revisions per type; % = percentage of total correct revisions with respect to total errors.

The L2 learners in the No-WCF condition revised a total of 10 gender assignment errors, with a range of 0 to 3 cases per text, and comprised the revision of 28% of all of their assignment errors produced during the revision sessions. The HL learners in the No-WCF group revised a total of 1 canonical gender assignment error, which corresponded to 33% of all of their errors. Since the L2 learners in the no-WCF group produced significantly more gender assignment errors than their HL counterparts ($p < 0.05$), a chi-square test was conducted to determine whether there was a significant difference in the percentage of total canonical gender assignment errors successfully revised by each group. The results revealed no significant difference in the percentage of canonical gender assignment errors correctly revised by the L2 and HL learners in the no-WCF group, $\chi^2 = 0.41$, $p = 0.52$.

Regarding the revision of gender agreement errors, the L2 learners in the no-WCF group provided the revision of 6 gender agreement errors, ranging from 0 to 1 error per text, which...
entailed 23% of all of their agreement errors; whereas their HL counterparts revised a total of 3
gender agreement revisions that involved 60% of all of their errors made during the treatment
period. The results of the second chi-square test revealed that the percentage of gender
agreement errors successfully revised by the HL learners in the No-WCF group was higher as
compared to their L2 learner counterparts, \( \chi^2 = 16.5, p = 0.00001, d = 0.49 \)

*Comparison of Revision of Canonical Gender Marking errors across groups (L2 and HL)*

The results of the independent samples t-tests that compared the total number of revisions
of non-canonical gender errors across the participants in the two group conditions showed that
the total average number of revisions of canonical gender assignment errors was significantly
higher in the L2 learners in the WCF group (M=0.37; SD=0.69) than in the L2 learners in the
No-WCF group (M=0.09; SD=0.4), \( t(161.7)= 3.5, p= 0.001, d = 0.496; \) and in the HL learners
(M=0.02; SD= 0.139) in the No-WCF group, \( t(116.26)= 4.9, p= 0.000, d = 0.496. \) Similarly, the
t-test results revealed that the L2 learners in the WCF group provided significantly more
revisions of canonical gender agreement errors (M=0.23; SD=0.52) than the L2 learners
(M=0.06; SD= 0.23) in the No-WCF group, \( t(137.9)= 2.9, p= 0.003 d = 0.42; \) and the HL
learners in the No-WCF group (M=0.06; SD=0.235), \( t(150.1)= 2.7, p= 0.007, d = 0.496. \)

A comparison between the HL learners in the WCF group and the HL and L2 learners in
the No-WCF group exhibited some differences in the revisions of canonical gender marking
errors. Given that the t-test results showed significant differences in the number of gender
assignment errors between the HL learners in the WCF group and the L2 learners in the No-
WCF group, a chi-square test was performed to compare the proportion of errors successfully
revised. The results revealed that the percentage of gender assignment errors correctly revised by
the HL learners was significantly higher compared to the L2 learners in the No-WCF group, \( \chi^2 \)
= 9.89, \( p = 0.002, d=0.64 \). Regarding the revision of gender agreement errors, since the average number of errors was similar between the HL learners in the WCF group and the L2 learners in the No-WCF group, an independent samples t-test was computed to compare the total number of revisions. The results revealed that, while the HL learners in the WCF group averaged more canonical gender agreement revisions (\( M=0.18; SD= 4.8 \)) than the L2 learners in the No-WCF group (\( M=0.06; SD= 0.23 \)), the differences were not statistically significant, \( t(61.6)= 1.7, p=0.095 \). Similarly, the t-tests results also showed that the total average number of canonical gender error revisions between the HL learners in the WCF group and their HL counterparts in the No-WCF group did not differ significantly, including both the number of revisions of canonical gender assignment, \( t(74.1)= 1.4, p= 0.169 \); and canonical gender agreement errors, \( t(72.6)= 1.59, p= 0.115 \).

In conclusion, the results indicate that the L2 learners who received grammar correction were able to revise significantly more canonical gender marking errors as compared to the L2 and HL learners who did revision without WCF. The HL learners in the treatment group, on the contrary, only provided comparatively more successful revisions of assignment errors than the L2 learners; while the other comparisons between the HL learners in the WCF and the L2 and HL learners in the No-WCF remained not significant.

**Comparison of Accuracy Development of Canonical Gender Marking (L2 and HL)**

A series of scatter plot diagrams were computed in order to observe the canonical gender marking accuracy trends followed by the L2 and HL learners in the two group conditions. The first diagram, displayed in Figure 4.24 below, shows that the L2 and HL learners in the WCF and No-WCF groups followed different directions in terms of canonical gender marking accuracy development.
As shown in Figure 4.24, the L2 learners in the WCF group experienced a slight increase in terms of gender assignment accuracy, while their HL counterparts in the WCF group underwent a minor decrease over time. With respect to the No-WCF condition, the results show that the accurate production of gender assignment in the L2 learners declined over time, whereas the HL learners consistently scored at ceiling or near ceiling.

The accuracy trends shown in Figure 4.25 below also reveal differences in the production of canonical gender agreement in the L2 and HL learners in the WCF and No-WCF groups.
The graph results indicate that the treatment did not have a positive effect on the accuracy development of canonical gender agreement of the participants in the WCF group, as the accurate production of the L2 learners declined, while the accuracy of the HL learners remained stable over time. On the other hand, the accuracy trend of the L2 learners in the no-WCF condition followed a stable trajectory over time, while their HL counterparts increased their accuracy over time and eventually reached ceiling.

4.3.2. Research question 2b: Non-Canonical Gender Marking

This section compares the mean production, error revision, accurate production and distribution of non-canonical gender marking between L2 and HL learners in the WCF and No-WCF groups.

*Comparison of Non-Canonical Gender Marking Production (L2 and HL)*

A comparison of the non-canonical gender marking production between the L2 and HL learners in the WCF group revealed that the number of cases and frequency of each canonical gender marking was comparable, as shown in 4.26 below.
The L2 learners in the WCF group produced a total of 701 cases of non-canonical gender assignment, with an average of 57 instances per text (SD= 3.5), while their HL counterparts in the WCF group totaled 401 cases of canonical gender assignment and averaged 6.27 cases per text (SD=3.8). The L2 learners in the WCF group also totaled 78 cases of non-canonical gender agreement, with an average of 0.64 instances per text (SD=1.1), while the HL learners in the same group produced 42 cases of canonical gender agreement and averaged roughly the same number of gender agreement with 0.66 cases per text (SD= 0.9). With respect to gender marking distribution for each type, the results in Figure 4.26 reveal a similar frequency of production in gender agreement in the L2 learners (90%) and the HL learners (90.5%) and gender agreement between the L2 learners (10%) and the HL learners (9.5%) in the WCF group.

Independent samples t-tests were computed to compare the total average production of non-canonical gender assignment and canonical gender agreement of the L2 and HL learners in the WCF group. The t-tests results showed no significant difference in the production of non-canonical gender assignment between the L2 learners (M=5.7; SD=3.5) and the HL learners (M=6.27; SD=3.8), $t(185)=-1.1, p = 0.309$. Similarly, the results revealed a comparable average...
production of non-canonical gender agreement between the L2 learners (M=0.64; SD=1.1) and the HL learners (M=0.66; SD=0.9) in the WCF group; \( t(183) = -0.75, p = 0.94 \).

The non-canonical gender marking production in the No-WCF group also revealed comparable results between the L2 and HL learners in terms of distribution with considerably more cases of gender assignment as compared to gender agreement, as shown in Figure 4.27.

Figure 4.27
*Distribution of Non-Canonical Gender Marking Production per Learner (No WCF Group)*

![Bar chart showing distribution of non-canonical gender marking](chart)

*Notes.* No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of gender marking cases; % = percentage of total gender marking

The L2 learners in the No-WCF group totaled 738 instances of non-canonical gender assignment, and averaged of 5.9 cases per text (SD= 3.4), whereas the HL learners in the No-WCF group produced a total of 432 cases of non-canonical gender assignment with a higher average of cases per text (M= 6.8; SD=3.7). Additionally, the L2 learners in the No-WCF group produced a total of 99 instances of non-canonical gender agreement, with an average of 0.79 cases per text (SD=1.1), whereas their HL counterparts produced a total of 61 cases of non-canonical gender agreement with a higher average of cases per text (M= 0.79; SD= 1.1). In terms of the distribution of non-canonical gender marking, similar results were found between the L2
and HL learners in the No-WCF group, including non-canonical gender assignment, (L2=88.2%; HL=87.6%) and gender agreement (L2=11.8%; HL=12.4%)

A second set of independent samples t-tests was computed to determine whether the mean production differences between the L2 and HL participants in the No-WCF group were significant. The results of the first independent samples t-test revealed that the L2 learners (M=5.86; SD=3.4) and the HL learners (6.75; SD=3.34) did not differ significantly in the total average production of non-canonical gender assignment \( t(188) = -1.71, p = 0.09 \). In addition, no significant differences were found in terms of non-canonical gender agreement between the L2 learners (M=0.79; SD=1.01) and the HL learners (0.97; SD=1.14), \( t(187) = -1.08, p = 0.28 \).

Comparison of Non-Canonical Gender Marking Errors within group (L2 and HL)

The results revealed that there were some differences in terms of error production and distribution between the L2 and HL learners in the WCF and No-WCF groups, as shown in Figure 4.28 and Figure 4.29.

Figure 4.28

Distribution of Non-Canonical Gender Marking Errors per learner type (WCF group)

Notes. WCF = revision with corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of gender marking errors; % = percentage of total gender marking errors per type.

The results in Figure 4.28 show that the L2 learners in the WCF group produced a total of 77 non-canonical gender assignment errors, with a range of 0 to 6 errors per text, and also 30
non-canonical gender agreement errors, ranging from 0 to 6 errors per text. The HL learners in the WCF group totaled 16 gender assignment errors, with a range of 0 to 2 errors per text, and 10 gender agreement errors, with a range from 0 to 2 errors. In addition, the results of the treatment group also revealed that assignment errors were more frequent in the L2 learners (72%) than in the HL learners (61.5%); while, agreement errors, in turn, were comparatively more frequent in the HL learners (38.5%) as compared to the L2 learners (28%).

A t-test comparison on the total average of non-canonical gender marking errors produced by the L2 and HL learners was calculated to determine whether there were differences in terms of overall accuracy. The results revealed that the L2 learners in the WCF group (M=0.63; SD=1.1) averaged significantly more non-canonical gender assignment errors than their HL counterparts (M=0.25; SD=0.64) in the WCF group, $t(183.6)= 2.9$, $p= 0.004$, $d= 0.42$ In contrast, no significant differences were found in the total average of non-canonical gender agreement errors between the L2 (M=0.25; SD=0.72) and HL learners (M=0.16; SD=0.48) in the WCF group, $t(184)= 0.89$, $p= 0.37$.

Regarding the non-canonical gender marking error production in the No-WCF group, the results revealed a similar distribution in the frequency of errors between the L2 and HL learners, as shown in Figure 4.29 below.

Figure 4.29
Distribution of Non-Canonical Gender Marking Errors per learner type (No WCF group)
Notes. No-WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of gender marking errors; % = percentage of total gender marking errors per type.

The L2 learners in the No-WCF produced 96 non-canonical gender assignment errors, which ranged from 0 to 6 errors per text, and also produced 27 non-canonical gender agreement errors, ranging from 0 to 2 errors per text. The HL learners in the No-WCF group produced a total of 7 assignment errors, with a lower range of 0 to 1 error per text, and 2 agreement errors, which ranged from 0 to 1 errors. The results in Figure 4.29 also indicate a similar distribution of gender marking errors between the L2 and HL learners in the No-WCF condition, including gender assignment (L2= 78%; HL= 77.8%) and agreement (L2= 22%; HL= 22.2%).

The t-test comparison revealed that, similar to the WCF group, the overall average number of gender assignment errors was significantly higher in the L2 learners (M=0.76; SD=1.17) than in the HL learners (M=0.11; SD=0.32), t(156.9)=5.9, p= 0.00002, d= 0.75 in the No-WCF group. Additionally, the t-test results also found that the L2 learners averaged significantly more non-canonical gender agreement errors (M=0.21; SD=0.5) than their HL counterparts (M=0.03; SD=0.18), t(177)= 3.86, p= 0.0001. d= 0.48.

Comparison of Non-Canonical Gender Marking Errors across groups (L2 and HL)

Two independent samples t-tests were performed to determine whether the L2 and HL learners across the two groups differed in the average number of non-canonical gender marking errors. First, a comparison between the L2 learners in the WCF group (M=0.63; SD=1.13) and the L2 learners in the No-WCF group (M=0.76; SD=1.17) showed no significant variation in the rate of non-canonical gender assignment errors, t(247)= -0.93, p= 0.36. Likewise, the total average number of canonical gender agreement errors was comparable between the L2 learners
in the WCF group (M=0.25; SD=0.72) and their L2 counterparts in the No-WCF group (M=0.21; SD=0.47), $t(246)= 0.412, p = 0.68$.

A second set of t-tests compared the total average number of non-canonical gender marking errors between the HL learners in the WCF and No-WCF groups. The results indicated that the HL learners in the WCF group averaged more non-canonical gender errors than their HL counterparts in the No-WCF group; however, the variation was not statistically significant in gender assignment, $t(91.57)= 1.57, p = 0.12$; and approached, but did not reach significance, in gender agreement, $t(79.6)= 1.96, p = 0.053$.

Comparison of Revision Non-Canonical Gender Marking errors within group (L2 and HL)

The results shown in Table 4.19 below reveal that the average number of non-canonical gender assignment corrections received by the L2 and HL learners in the WCF condition was comparable.

<table>
<thead>
<tr>
<th>Learner</th>
<th>WCF Group</th>
<th>No WCF Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 (n=12)</td>
<td>N= 38 (M=1.52; SD=1.5)</td>
<td>N= 0 (M=0)</td>
</tr>
<tr>
<td>HL (n=6)</td>
<td>N= 26 (M=1.30; SD= 1.1)</td>
<td>N= 0 (M=0)</td>
</tr>
</tbody>
</table>

Notes. Corrections = Total number of gender assignment corrections; Mean = average number of corrections per participant; WCF = revision with corrective feedback; No WCF = revision without corrective feedback.

The L2 learners in the WCF group received a total of 38 non-canonical gender assignment corrections, ranging from 0 to 6 corrections per text, while their HL counterparts received a total of 26 error assignment corrections, with a range of 0 to 3 errors per text. Furthermore, a t-test comparison revealed that the differences in the mean number of corrections
that the L2 learners (M=1.52; SD=1.53) and HL learners (M=1.30; SD=1.13) in the WCF group received during the treatment period were not statistically significant, $t(43)= 0.54, p= 0.595$.

With respect to the number of non-canonical gender agreement corrections, the results in table 4.20 below show that the L2 learners averaged more corrections than the HL learners in the WCF group.

Table 4.20

<table>
<thead>
<tr>
<th>Learner</th>
<th>WCF Group</th>
<th>No WCF Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L2 (n=12)</td>
<td>HL (n=6)</td>
</tr>
<tr>
<td></td>
<td>N= 20</td>
<td>(M= 0.80; SD = 0.7)</td>
</tr>
<tr>
<td></td>
<td>(M= 0.50; SD = 0.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N= 10</td>
<td>N= 0</td>
</tr>
<tr>
<td></td>
<td>N= 0</td>
<td>N= 0</td>
</tr>
</tbody>
</table>

Notes. Corrections = Total number of gender agreement corrections; Mean = average number of corrections per participant; WCF = revision with corrective feedback; No WCF = revision without corrective feedback.

The L2 learners in the WCF group received a total of 20 non-canonical gender agreement corrections, which ranged from 0 to 2 corrections per text, while the HL learners totaled 10 non-canonical gender agreement corrections, ranging from 0 to 2 errors per text. A mean comparison of the average number of corrections of canonical gender agreement showed no significant differences between the L2 learners in terms of errors (M=0.8; SD=0.71) and their HL counterparts (M=0.5; SD=0.69) in the WCF group, $t(43)= 1.43, p= 0.16$.

With respect to the revision of non-canonical gender marking errors, there were differences between the two group conditions in terms of total number of revisions and distribution, as shown in Figure 4.30 and Figure 4.31 below.
The L2 learners in the WCF group revised a total of 33 errors non-canonical gender assignment errors, ranging from 0 to 6 revisions per text, and comprised the revision of 72% of all of their non-canonical gender assignment errors produced during the treatment period. The HL learners in the WCF group, in turn, provided revisions of 18 errors of non-canonical gender assignment errors with a range of 0 to 3 revisions per text, which totaled to 55% of all of their errors made during the treatment period. Since the L2 and HL learners in the WCF group differed significantly in the total number of non-canonical gender assignment errors ($p < 0.05$), a chi-square test was computed to examine whether differences in the percentage of successful revisions were statistically significant between the participants. The results revealed no significant difference between the L2 and HL in the percentage of total non-canonical gender assignment errors successfully revised, $\chi^2 = 2.27, p = 0.13$.

With respect to non-canonical gender agreement, the L2 learners in the WCF group revised a total of 13 errors, which consisted of 63% of all of their gender agreement errors, with
a range of 0 to 2 revisions per text. The HL learners in the WCF group provided a total of 8 gender agreement revisions, ranging from 0 to 2 revisions per text, which, in turn, represented 67% of all of their errors. A t-test comparison of the total number of successful gender agreement revisions showed no differences between L2 and HL learners in the WCF group, \( t(43)= 0.57, p= 0.57. \)

Regarding the No-WCF condition, the proportion of non-canonical gender marking errors revised by the L2 learners was higher as compared to their HL counterparts, as shown in Figure 4.31 below.

Figure 4.31
Total Number and Percentage of Non-Canonical Gender Marking Revisions with respect to Total Errors (L2 and HL)

![Graph showing the total number and percentage of non-canonical gender marking revisions with respect to total errors for L2 (No-WCF) and HL (No-WCF).]

Notes. No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of correct gender marking revisions per type; % = percentage of total correct revisions with respect to total errors

The L2 learners in the No-WCF group revised a total of 16 non-canonical gender assignment errors, with a number of revisions that ranged from 0 to 3 per text, which involved the revision of 72% of all of their assignment errors. The HL learners in the No-WCF group revised 1 error of non-canonical gender assignment, which corresponded to the revision of 14% of all of their assignment errors. Given the significant differences in terms of number of errors, a
chi-square test was calculated to determine whether the L2 and HL learners differed in terms of percentage of successful revisions with respect to their total errors. The results revealed no significant differences between the two learners, $\chi^2 = 0.76$, $p = 0.38$.

In terms non-canonical gender agreement revisions, the L2 learners revised a total of 13 errors, which entailed 63% of all of their agreement errors, ranging from 0 to 1 correct revision per text; while the HL learners did not provide any gender agreement error revisions. The chi-square test results revealed that the percentage of successful revisions was higher in the L2 learners as compared to their HL counterparts in the No-WCF group, $\chi^2 = 10$, $p = 0.0016$, $d=0.63$

Comparison of Revision Non-Canonical Gender Marking errors across groups (L2 and HL)

A mean comparison between the L2 and HL learners across the two group conditions showed differences in terms of the mean number of gender marking revisions. The results of the independent samples t-tests showed that the L2 learners that received the WCF treatment averaged significantly more revisions of non-canonical gender assignment errors ($M=1.32$; $SD=1.46$) than the L2 learners in the No-WCF condition ($M=0.32$; $SD=0.62$), $t(28.4)=3.27$, $p=0.003$, $d=0.89$; and the HL learners ($M=0.13$; $SD=0.35$) in the No-WCF condition, $t(30.1)=3.7$, $p=0.001$, $d=1.03$. Likewise, the t-test results indicated that the L2 learners in the WCF group were able to revise significantly more canonical gender agreement errors ($M=0.52$; $SD=0.71$) than the L2 learners ($M=0.10$; $SD=0.30$) in the No-WCF group, $t(28.4)=2.8$, $p=0.009$, $d=0.77$; and the HL learners in the No-WCF group ($M=0.0$; $SD=0.0$), $t(24)=3.6$, $p=0.001$ $d=1.03$.

Furthermore, a series of t-test comparisons also found that the group of HL learners who received WCF ($M=0.90$; $SD=0.9$) averaged significantly more non-canonical gender assignment revisions than both the L2 learners in the No-WCF group ($M=0.32$; $SD=0.62$), $t(26.3)=2.61$, $p=
0.015, \( d = 0.75 \); and the HL learners (M=0.13; SD= 0.35) in the No-WCF group, \( t(25.9) = 3.24, p= 0.003, d= 1.13 \). The t-test results on non-canonical gender agreement error revision also revealed significant differences in the total average number of revision between the HL learners in the WCF group (M=0.40; SD= 0.68) and the HL learners in the No-WCF group (M=0.0; SD= 0.0), \( t(19.0)= 2.62, p= 0.017 \ d= 0.83 \); and a near-significant difference when compared to the total number of revisions made by the L2 learners (M=0.10; SD= 0.30) in the No-WCF group, \( t(22.1)= 1.9, p= 0.07, d= 0.57 \).

**Comparison of Accurate Production of Non-Canonical Gender Marking Forms after Revision (L2 and HL)**

The descriptive results in Table 4.21 show the total number of traceable non-canonical gender marking forms by type of learner (L2 and HL) and group condition. All of the forms in the count corresponded to gender assignment errors (N= 43) given that no traceable instances of gender agreement forms produced by the same participant were found in the dataset.

<table>
<thead>
<tr>
<th>Learner</th>
<th>WCF Group</th>
<th>No WCF Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 (n=7)</td>
<td>L2 (n=5)</td>
<td>L2 (n=5)</td>
</tr>
<tr>
<td>HL (n=1)</td>
<td>HL (n=1)</td>
<td>HL (n=1)</td>
</tr>
<tr>
<td>Cases</td>
<td>N= 16</td>
<td>N= 4</td>
</tr>
<tr>
<td></td>
<td>N= 22</td>
<td>N= 1</td>
</tr>
</tbody>
</table>

**Notes.** WCF = revision with corrective feedback; No WCF = revision without corrective feedback; n = participants per group; Cases: individual cases of non-canonical gender assignment; Mean = average number of cases per participant; SD = standard deviation

A total of 16 traceable forms were found in the production of L2 learners and 4 forms in the production of the HL learners in the WCF group. On the other hand, a total of 22 traceable forms were found in the production of the L2 learners in the No-WCF group, while only 1 form
was found in their HL counterpart in the No-WCF group. All forms were classified according to whether the same revised form was (a) immediately produced by the same learner in the following text, (b) produced by the same learner in a later but not immediate text, or (c) produced in two (or more) different texts.

Descriptive statistics on the effects of error revision on the accurate production of non-canonical gender marking forms in subsequent production are shown in Table 4.22 below.

<table>
<thead>
<tr>
<th>Learner</th>
<th>WCF Group</th>
<th>No WCF Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n=8)</td>
<td>(n=6)</td>
</tr>
<tr>
<td></td>
<td>L2 (N=7)</td>
<td>L2 (N=5)</td>
</tr>
<tr>
<td></td>
<td>HL (N=1)</td>
<td>HL (N=1)</td>
</tr>
<tr>
<td>Revision + Immediate Production</td>
<td>80% (N=5)</td>
<td>100% (N=1)</td>
</tr>
<tr>
<td></td>
<td>- (N=0)</td>
<td>- (N=0)</td>
</tr>
<tr>
<td>Revision + Later Production</td>
<td>10.7% (N=5)</td>
<td>100% (N=1)</td>
</tr>
<tr>
<td></td>
<td>9.4% (N=4)</td>
<td>- (N=0)</td>
</tr>
<tr>
<td>Revision + Production in two different texts</td>
<td>3.6% (N=6)</td>
<td>0% (N=2)</td>
</tr>
<tr>
<td></td>
<td>0% (N=1)</td>
<td>- (N=0)</td>
</tr>
<tr>
<td>No revision + Immediate Production</td>
<td>- (N=0)</td>
<td>- (N=0)</td>
</tr>
<tr>
<td></td>
<td>60% (N=5)</td>
<td>- (N=0)</td>
</tr>
<tr>
<td>No revision + Later Production</td>
<td>- (N=0)</td>
<td>- (N=0)</td>
</tr>
<tr>
<td></td>
<td>16.7% (N=12)</td>
<td>0% (N=1)</td>
</tr>
<tr>
<td>No Revision + Production in two different texts</td>
<td>- (N=0)</td>
<td>- (N=0)</td>
</tr>
<tr>
<td></td>
<td>- (N=0)</td>
<td>- (N=0)</td>
</tr>
</tbody>
</table>

*Notes.* % = Percentage of correct use of non-canonical gender assignment forms; N = number of non-canonical gender agreement forms; Revision = a non-canonical gender assignment form was corrected during the revision session; No Revision = the correct non-canonical gender assignment was not provided during the revision session; Immediate Production = the same non-canonical gender assignment form was used after the revision stage; Production in two different texts = the same non-canonical gender assignment form was used in two different texts; Later Production = the same non-canonical gender assignment form was used in a subsequent, but not immediate, text.

Given the reduced number of traceable gender assignment forms found in the production of both the HL learners in the WCF group (N=4) and the HL learners in the No-WCF group
(N=1), only the data of the L2 learners in the WCF and No-WCF was included in the final comparison, shown in Figure 4.32 below.

The results show that the accurate production of non-canonical gender assignment forms in the L2 learners in both groups was the higher when a revised form was immediately used in production (80%) as compared to no revision (60%). On the other hand, the accurate use of revised non-canonical gender assignment forms decreased drastically when the same form was used in later production (WCF=11%; No WCF= 9%). Lastly, the results revealed that accurate production was the lowest in both the L2 learners in the WCF group (4%) and the L2 learners in the No-WCF group (0%) when the same revised non-canonical gender assignment form was produced in more than one text.

4.3.3. Research question 2c: Definite Articles in Obligatory Contexts

The production, revision and accuracy development of definite articles in obligatory contexts were compared within and across the WCF and No-WCF groups in order to assess any differences between the L2 and HL learners in the two conditions.
Production of Definite Articles in Obligatory Contexts (L2 and HL)

The results, shown in 4.23 below, revealed comparable results between the L2 and HL learners in the WCF and No-WCF groups in the mean and total number of definite articles in obligatory contexts.

Table 4.23
Total errors of definite article omission in obligatory contexts (L2 and HL)

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (M)</th>
<th>Total Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>L2 (WCF)</td>
<td>4.41</td>
<td>582</td>
</tr>
<tr>
<td>HL (WCF)</td>
<td>4.7</td>
<td>310</td>
</tr>
<tr>
<td>L2 (No WCF)</td>
<td>4.44</td>
<td>590</td>
</tr>
<tr>
<td>HL (No WCF)</td>
<td>4.48</td>
<td>291</td>
</tr>
</tbody>
</table>

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; M = mean number of definite articles in obligatory contexts

A comparison between the participants in the two conditions revealed a similar rate of definite articles in obligatory contexts in the overall production with an average of 4.41 cases per text (SD=3.3) in the L2 learners in the WCF group, and an average of 4.7 cases in the HL learners in the WCF group (SD= 3.0). Similarly, an average of 4.44 cases (SD=2.8) of definite articles in obligatory contexts were found in the production of the L2 learners in the No-WCF group, while their HL counterparts averaged a total of 4.48 cases (SD=2.5).

Comparison of Definite Article Omission Errors (L2 and HL)

The results indicate that the L2 learners in the WCF group produced fewer errors of
article omission than their L2 counterparts in the No-WCF group, whereas the total number of errors in the HL learners was similar in the two conditions, as shown in Figure 4.33 below.

Figure 4.33
*Total errors of definite article omission in obligatory contexts (WCF and No-WCF)*

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; M = mean number of errors of definite article omission in obligatory contexts

The L2 learners in the WCF group produced a total of 70 errors of definite article omission, with an average of 0.53 errors per text (SD= 1.14), while their L2 counterparts in the No-WCF group produced a total of 144 errors, and averaged 1.07 errors per text (SD=1.73). In addition, the HL learners in the WCF group averaged 0.39 article omission errors (SD 1.08), whereas the HL learners in the No-WCF group averaged 0.66 article omission (SD=1.28),

A within group comparison between the L2 and HL learners in the WCF condition revealed no significant differences, $t(195)= 0.83, p = 0.40$. With respect to the No-WCF group, the HL learners averaged fewer article omission errors than their L2 counterparts (M=1.08; SD=1.73), which approached, but did not reach significance, $t(165.6)= 1.86, p = 0.06$.
A comparison between the L2 and HL learners across groups revealed some differences in terms of the number of definite article omission errors. The t-test results indicated that the L2 learners in the No-WCF group produced significantly more definite article omission errors (M=1.07; SD=1.72) than their L2 counterparts in the WCF group, \( t(228.1)= -3.03, p = 0.003 \), with a small size effect, \( d=0.18 \). The results also reveal that the HL learners in the WCF and No-WCF groups, on the contrary, did not differ significantly in the average number of article omission errors, \( t(129)= -1.3, p = 0.207 \).

Comparison of Revision of Definite Article Omission errors within groups (L2 and HL)

The results reveal that, as part of the treatment, the L2 learners received more definite article omission revisions than their HL counterparts in the WCF group; whereas the L2 and HL learners in the control condition did not receive any corrections on their errors, as shown in Table 4.24 below.

<table>
<thead>
<tr>
<th>Learner</th>
<th>WCF Group</th>
<th>No WCF Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L2 (n=12)</td>
<td>L2 (n=12)</td>
</tr>
<tr>
<td></td>
<td>HL (n=6)</td>
<td>HL (n=6)</td>
</tr>
<tr>
<td>Corr.</td>
<td>N= 64</td>
<td>N=0</td>
</tr>
<tr>
<td></td>
<td>(M=0.67; SD=1.2)</td>
<td>(M=0)</td>
</tr>
</tbody>
</table>

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N= total number of corrections; M = mean of corrections provided

The L2 learners in the WCF group received a total of 64 corrections during the treatment period, ranging from 0 to 6 errors per text, and a mean of 0.67 corrections per text (SD=1.2). The HL learners in the WCF group received a total of 18 corrections, which ranged from 0 to 5
corrections per text, and averaged 0.34 corrections per text (SD=0.88). A t-test comparison revealed that the L2 learners averaged more corrections than their HL counterparts in the WCF group with differences that approached significance, \( t(135.5) = 1.9, p = 0.054 \).

Regarding the total number and average of revisions of definite article omission errors, the results, displayed in Figure 4.34, show differences between the L2 and HL learners within and across the two conditions.

Figure 4.34
*Total Number and Percentage of Article Omission Revisions with respect to Total Errors (L2 and HL Groups)*

<table>
<thead>
<tr>
<th></th>
<th>L2 (WCF)</th>
<th>HL (WCF)</th>
<th>L2 (No WCF)</th>
<th>HL (No WCF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>56</td>
<td>17</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>%</td>
<td>87.5%</td>
<td>94.4%</td>
<td>5.2%</td>
<td>8.6%</td>
</tr>
</tbody>
</table>

*Notes.* WCF = revision with corrective feedback; No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; \( N \) = number of correct article omission revisions per type; \% = percentage of total correct revisions with respect to total errors.

The L2 learners provided a total of 56 correct revisions of definite article omission errors, which ranged from 0 to 6 revisions per text, with a mean of 0.59 revisions, and involved the revisions of 87.5\% of all their errors of omission. The HL learners in the WCF group revised a total of 17 article omission errors, ranging from 0 to 5 revisions per text with an average of 0.32 cases (SD=0.87), which comprised the revision of 94.4\% of all of their errors produced during
the treatment period. A t-test comparison between the L2 and HL learners in the WCF condition found no significant differences in the total number of revisions, $t(146)= 1.54, p = 0.125$.

With respect to the No-WCF group, the L2 learners revised a total of 5 errors of article omission, ranging from 0 to 1 revision with an average of 0.11 instances per text (SD=0.32), which corresponded to the revision of 5.2% of all of the errors produced during the revision sessions. The HL learners in the No-WCF group provided a total of 3 correct revisions of omission errors, which ranged from 0 to 1 revisions, with an average of 0.16 per text (SD=0.50) that, in turn, involved 8.6% of all the errors. A mean comparison between the L2 and HL learners in the No-WCF group revealed comparable results in terms of revision of definite article omission errors, $t(63)= -4.8, p = 0.634$.

**Comparison of Revision Definite Article Omission errors across groups (L2 and HL)**

A comparison between the L2 and HL learners across groups showed some differences regarding the number of revisions of definite article omission errors. First, since the differences in the total number of article omission errors between the L2 learners in the WCF group and the L2 learners in the No-WCF group was significant ($p =< 0.05$), a chi-square test was performed to determine whether there was a significant difference between the percentage of errors revised with respect to the total number of errors made by each group. The results showed that the percentage of errors successfully revised by the L2 learners in the WCF group was significantly higher than their L2 counterparts in the No-WCF, $\chi^2 = 73.07, p = 0.00, d= 0.23$. With respect to the comparison between the L2 learners in the WCF group and the HL learners in the No-WCF group, given that there were not significant differences in terms of total article omission errors, an independent samples t-test was computed to compare the total number of revisions. The t-test results revealed that the L2 learners in the WCF averaged significantly more correct revisions of
article omission than the HL learners in the No-WCF group, with equal variances not assumed, \( t(57.9) = 2.69, p = 0.009, d=0.51 \).

A comparison between the HL in the WCF group with the L2 and HL learners in the No-WCF group revealed some differences in terms of revision of article omission errors. A t-test comparison revealed that, while the HL learners in the WCF group averaged more correct revisions (M=0.32; SD=0.87) than their HL counterparts in the No-WCF group (M=0.16; SD=0.50), the differences were not statistically significant, \( t(70) = 0.77, p = 0.445 \).

Furthermore, given that there were significant differences in the number of article omission errors between the HL learners in the WCF group and the L2 learners in the No-WCF group \( (p =< 0.05) \), a chi-square test was calculated to compare the percentage of errors revised with respect to the total number of article omission errors. The chi-square results showed that the percentage of errors successfully revised by the HL learners in the WCF group was significantly higher than the L2 learners in the No-WCF, \( \chi^2 = 79.89, p = 0.00, d= 0.22 \).

*Comparison of Accuracy Development of Definite Articles in Obligatory Contexts (L2 and HL)*

The results of the scatter plot diagram, displayed in Figure 4.35 below, shows differences between the trajectories followed by the L2 and HL learners in the WCF group and their counterparts in the No-WCF group with respect to the accurate production of definite articles in obligatory contexts.
The graph results indicate that the L2 and HL learners in the WCF group underwent similar gains in terms of the accurate production of definite articles in obligatory contexts, as shown by their accuracy trends. In contrast, the accurate production of the L2 learners in the no-WCF group declined slightly over time, whereas their HL counterparts experienced a moderate increase in accuracy.

4.3.4. Research question 2b: Present Subjunctive

Comparison of Production of Present Subjunctive (L2 and HL)

A comparison between the L2 and HL learners in the WCF and No-WCF groups revealed some similarities and differences in terms of distribution and contexts that triggered the use of the present subjunctive, as shown in Figure 4.36 below.
Regarding the L2 learners in the WCF group, the most frequent structure that required the use of the present subjunctive was elicited by verbs of denial and negation (N=26; 33%) followed by impersonal structures (N=18; 23%), verbs of influence and volition (N=6; 8%), adverbial clauses (N=5; 6%) and unknown antecedents in adjectival clauses (N=5; 6%). Similarly, the production of the subjunctive in the L2 learners in the No-WCF group was primarily triggered by verbs of denial and negation (N=18; 35%) and impersonal structures (N=14; 27%), followed by verbs of influence and volition (N=4; 8%), adverbial clauses (N=4; 8%) and unknown antecedents in adjectival clauses (N=4; 8%).
The results also showed similarities between the HL learners in the WCF and No-WCF groups with respect to the distribution of the structures that triggered the present subjunctive. The most frequent forms that elicited the use of the present subjunctive among the HL learners in the WCF group corresponded to unknown antecedents (N=9; 36%), impersonal structures (N=7; 28%), verbs of denial and negation (N=4; 16%), adverbial clauses (N=3; 12%) and verbs of influence and volition (N=2; 8%). Likewise, the occurrence of the present subjunctive in the HL learners in the No-WCF group was predominantly related to verbs of denial and negation (N=15; 37.5%) and unknown antecedents (N=12; 30%), followed by impersonal structures (N=5; 12.7%), adverbial clauses (N=4; 10%) and verbs of influence and volition (N=1; 2.5%).

Comparison of Present Subjunctive Errors within Group (L2 and HL)

Two sets of independent sample t-tests were conducted to compare the total number of subjunctive errors produced by the L2 and HL in each of the two conditions. The t-test results showed that there were fewer subjunctive errors in the written production of the HL learners in the WCF group (M=0.36; SD=0.49) compared to the L2 learners in the same condition (M=0.87; SD=0.34), \(t(31.7)=4.8, p=0.00003\) (equal variances not assumed), \(d=1.14\).

Similarly, the results of the second t-test also revealed that the HL learners in the No-WCF group (M=0.51; SD=0.51) produced significantly fewer errors than their L2 counterparts (M=0.86; SD=0.35), \(t(64.7)=3.65, p=0.001\) (equal variances not assumed), \(d=0.8\).

With respect to specific forms and structures that elicited the use of the subjunctive, the descriptive results in Table 4.25 show that comparatively the L2 learners in the WCF and No-WCF produced more present subjunctive errors and were less accurate than their HL counterparts in the majority of the categories.
Table 4.25
*Total errors and percentage accurate use of Present Subjunctive per type and learner*

<table>
<thead>
<tr>
<th>Category</th>
<th>WCF</th>
<th></th>
<th>No-WCF</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L2</td>
<td>HL</td>
<td>L2</td>
<td>HL</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Accuracy</td>
<td>N</td>
<td>Accuracy</td>
</tr>
<tr>
<td>Influence/volition</td>
<td>6</td>
<td>0%</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Negation</td>
<td>21</td>
<td>19.2%</td>
<td>4</td>
<td>0%</td>
</tr>
<tr>
<td>Opinion</td>
<td>10</td>
<td>16.7%</td>
<td>1</td>
<td>87.5%</td>
</tr>
<tr>
<td>Adverbial</td>
<td>2</td>
<td>50%</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Adjectival</td>
<td>3</td>
<td>0%</td>
<td>3</td>
<td>40%</td>
</tr>
<tr>
<td>Doubt</td>
<td>1</td>
<td>0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possibility</td>
<td>5</td>
<td>16.7%</td>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>47</strong></td>
<td><strong>17.5%</strong></td>
<td><strong>9</strong></td>
<td><strong>62.5%</strong></td>
</tr>
</tbody>
</table>

*Notes.* WCF = revision with corrective feedback; No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of present subjunctive errors per type; % = percentage of accurate production

The results indicate that the L2 learners in the WCF group totaled 47 errors of present subjunctive, which involved 17.5% of accuracy, while their HL counterparts produced a total of 9 errors with 62.5% accuracy. The HL learners were more accurate in 4 out of 6 categories, including verbs of influence and volition, L2 learners (0%), HL learners (100%); impersonal structures expressing opinion, L2 learners (16.7%), HL learners (87.5%); adverbial clauses, L2 learners (50%), HL learners (100%); adjectival clauses with an unknown antecedent, L2 learners (0%), HL learners (40%). The overall use of the present subjunctive elicited by negated verbs in the main clause was more accurate in the L2 learners (19.2%) than in the HL learners (0%) in the
The WCF group in addition to impersonal structures that express possibility, L2 learners (17.5%), HL learners (0%).

With respect to the accurate use and frequency of present subjunctive errors among the participants in the No-WCF group, the results in Table 4.25 show that the L2 learners produced a total of 35 errors with 18.3% accuracy, whereas their HL counterparts totaled 17 errors, with 59% accuracy.

The HL learners in the No-WCF group were more accurate than their L2 counterparts in all 6 categories, including verbs of influence and volition, L2 learners (0%), HL learners (100%); negated verbs in the main clause, L2 learners (5.6%), HL learners (21.4%), impersonal structures expressing opinion, L2 learners (22.2%), HL learners (25%); adverbial clauses, L2 learners (75%), HL learners (100%); adjectival clauses with an unknown antecedent, L2 learners (0%), HL learners (66.7%) and impersonal structures that express possibility, L2 learners (25%), HL learners (100%).

Additionally, the results revealed that the L2 learners in the WCF and no-WCF groups produced more errors of oversuppliance of the present subjunctive than their HL counterparts, as shown in Table 4.26 below.

Table 4.26
Total errors of Oversuppliance of Present Subjunctive

<table>
<thead>
<tr>
<th>Category</th>
<th>WCF</th>
<th>No-WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oversuppliance of Subjunctive</td>
<td>N=20</td>
<td>N=0</td>
</tr>
<tr>
<td></td>
<td>N=8</td>
<td>N=3</td>
</tr>
</tbody>
</table>

*Notes.* WCF = revision with corrective feedback; No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of overgeneralized use of subjunctive
The L2 learners in the WCF group produced a total of 20 errors of oversuppliance of the subjunctive, whereas the HL learners in the WCF group did not produce any instances of oversuppliance of the subjunctive. The results of the No-WCF group revealed a total of 8 errors, of oversuppliance of the present subjunctive whereas their HL counterparts produced a total of 3 cases of use of the present subjunctive in contexts where the present indicative was required.

Comparison of Present Subjunctive Errors across Groups (L2 and HL)

Two sets of independent sample t-tests were calculated to compare whether the L2 and HL differed in the total number of subjunctive errors across the two conditions. The results of the first t-test results revealed that the that there were fewer subjunctive errors in the written production of the HL learners in the No-WCF group (M=0.51; SD=0.51) compared to the L2 learners in the WCF group (M=0.87; SD=0.34), t(55.7)= 3.98, p= 0.0002 (equal variances not assumed), d=0.83.

The results of the second t-test also revealed that the HL learners in the WCF group (M=0.36; SD=0.49) produced significantly fewer subjunctive errors than the L2 learners in the No-WCF group (M=0.86; SD=0.35), t(36.6)= -4.56, p= 0.00006 (equal variances not assumed), d=1.17. Furthermore, a comparison between the L2 and HL learners across the two conditions revealed no significant differences in the total number of subjunctive errors between the L2 learners in the WCF and No-WCF groups, and also between the HL learners in the WCF and the HL learners in the No-WCF groups (p > 0.05).

In terms of the categories that elicited the use of the present subjunctive, a comparison between the L2 and HL learners across groups revealed that the overall accuracy rate of the present subjunctive across all sessions was highest in the HL learners in the WCF group (62.5%), followed by the HL learners in the No-WCF group (59%). In addition, the total accurate
production of the present subjunctive in the L2 learners (17.5%) in the WCF group was similar to their L2 counterparts in the No-WCF group (18.3%), as shown in Figure 4.36.

The descriptive results shown earlier in Table 4.25 revealed no instances of correct use of the present subjunctive triggered by verbs of influence/volition and in adjectival clauses with unknown antecedent among the L2 learners in the WCF and No-WCF groups. Furthermore, the highest overall accurate use of the subjunctive for all the learners was found in adverbial clauses, HL (100%), L2 (50%) learners in WCF group; HL (100%) and L2 (75%) learners in the No-WCF group, followed by the use of the subjunctive elicited by impersonal structures that express opinion; HL (87.5%), L2 (16.7%) learners in WCF group; HL (25%) and L2 (22.2%) learners in the No-WCF group. Furthermore, the frequency of errors of oversuppliance of the present subjunctive was the highest among the L2 learners in the WCF group, which totaled 29.8% of all errors (N=20), followed by the L2 learners in the No-WCF group with 18.6% of all errors (N=8), and the HL learners in the No-WCF group with 15% (N=3).

Comparison of Revision of Present Subjunctive errors within groups (L2 and HL)

The results displayed in Table 4.27 below show the differences in the average number of corrections of present subjunctive errors that the L2 and HL learners received during the treatment period.

<table>
<thead>
<tr>
<th>Table 4.27</th>
<th>Summary of corrections of present subjunctive errors per group and learner (L2 and HL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCF Group</td>
<td></td>
</tr>
<tr>
<td>Learner</td>
<td>L2 (n=12)</td>
</tr>
<tr>
<td>Corrections</td>
<td>N= 45</td>
</tr>
<tr>
<td>No WCF Group</td>
<td></td>
</tr>
<tr>
<td>Learner</td>
<td>L2 (n=12)</td>
</tr>
</tbody>
</table>
| Corrections | N= 0                                      | N= 0                                      | N= 0}
Notes. Corrections = Total number of gender agreement corrections; Mean = average number of corrections per participant; WCF = revision with corrective feedback; No WCF = revision without corrective feedback

The results revealed that the L2 learners in the WCF group received a total of 45 corrections of present subjunctive errors (SD=0.48), which ranged from 0 to 1 error per text with an average of 0.76 corrections per text. The HL learners in the WCF received a total of 3 corrections, ranging from 0 to 1 per text, and averaged 0.37 corrections per text (SD= 0.52) during the treatment period. A t-test comparison revealed that the L2 learners received significantly more corrections than the HL learners in the WCF group, \( t(63)= 2.59, p = 0.012 \).

In terms of the total number and percentage of revisions of present subjunctive errors, the proportion of errors revised by the L2 and HL learners in the WCF group was higher compared to the L2 and HL learners in the No-WCF group, as shown in Figure 4.37 below.

Figure 4.37
Total Number and Percentage of Revisions of Present Subjunctive with respect to Total Errors (L2 and HL Groups)

<table>
<thead>
<tr>
<th></th>
<th>L2 (WCF)</th>
<th>HL (WCF)</th>
<th>L2 (No WCF)</th>
<th>HL (No WCF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>39</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Revision Rate</td>
<td>56%</td>
<td>22%</td>
<td>7%</td>
<td>10%</td>
</tr>
</tbody>
</table>

Notes. WCF = revision with corrective feedback; No WCF = revision without corrective feedback; L2 = second language learner; HL = heritage language learner; N = number of correct
present subjunctive revisions per type; % = percentage of total correct revisions with respect to total errors

The L2 learners in the WCF group totaled 39 revisions of subjunctive errors including errors of oversuppliance, which comprised 58.2% of all of their errors produced during the revision sessions, ranging from 0 to 1 revisions per text, with an average of 0.68 revisions (SD=0.47). Additionally, the HL learners in the WCF group revised a total of 2 present subjunctive errors, which involved the revision of 22% of all of their errors, with a range of 0 to 1 revisions and an average of 0.25 revisions per text (SD=0.46). The results of a chi-square test revealed that the percentage of correct revisions was significantly higher in the L2 as compared to the HL learners in the WCF group, \( \chi^2 = 16.3, p = 0.00005, d = 0.13 \).

The results of the No-WCF group indicate that the L2 learners revised a total of 3 errors of present subjunctive, ranging from 0 to 1 revisions, and averaged 0.08 revisions per text (SD=0.27), which corresponded to the revision of 7.9% of all of their errors including errors of oversuppliance. The HL learners in the No-WCF group provided the correct revision of 2 errors, ranging from 0 to 1 per text, with an average of 0.11 revisions (SD=0.32), which entailed the revision of 10.5% of all of their errors. A chi-square test revealed that the percentage of revisions with respect to the total number of present subjunctive errors was similar between the L2 and HL learners in the No-WCF group, \( \chi^2 = 0.367, p = 0.54 \).

**Comparison of Revision of Present Subjunctive errors across groups (L2 and HL)**

Since the L2 and HL learners in the WCF and No-WCF groups differed in the total number of subjunctive errors, a series of chi-square tests were calculated to determine whether there were significant differences in the percentages of total present subjunctive errors that were successfully revised by each group.
The chi-square results revealed that the proportion of present subjunctive errors correctly revised by the L2 learners in the WCF group was higher than both the L2 learners in the No-WCF group, \( \chi^2 = 38.2, p = 0.00, d = 0.44 \), and the HL learners in the No-WCF group, \( \chi^2 = 33.68, p = 0.00, d = 0.33 \). Likewise, the results showed that the percentage of successfully revised errors was significantly higher in the HL learners in the WCF condition as compared to the L2 learners in the No-WCF group, \( \chi^2 = 6.64, p = 0.01, d = 0.22 \), and the HL learners in the No-WCF group, \( \chi^2 = 4.06, p = 0.04, d = 0.14 \).

**Comparison of Accurate Production of Specific Present Subjunctive Forms (L2 and HL)**

The descriptive results displayed in Table 4.28 show the number of traceable forms that required the use of the present subjunctive.

<table>
<thead>
<tr>
<th>Learner</th>
<th>WCF Group</th>
<th>No WCF Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L2 (n=7)</td>
<td>HL (n=1)</td>
</tr>
<tr>
<td>Cases</td>
<td>N=12</td>
<td>N=1</td>
</tr>
<tr>
<td></td>
<td>L2 (n=8)</td>
<td>HL (n=3)</td>
</tr>
<tr>
<td></td>
<td>N=16</td>
<td>N=5</td>
</tr>
</tbody>
</table>

**Notes.** WCF = revision with corrective feedback; No WCF = revision without corrective feedback; Learners = Second and Heritage Language Learners; Cases: individual cases of non-canonical gender assignment; Mean = average number of cases per participant; SD = standard deviation.

A total of 12 traceable forms that required the use of the present subjunctive were extracted from the written data of seven L2 learners in the WCF group, and only one traceable form was found in the written data of the HL learners in the WCF group. Additionally, a total of 16 traceable forms that required the use of the present subjunctive were located in the production of L2 learners in the No-WCF group (N=8), whereas 5 forms were found in the written production of the HL learners in the No-WCF group (N=3).
All forms were further classified according to whether (a) the same form that required the use of the subjunctive was immediately produced by the same learner in the following text, (b) the same form was produced by the same learner in a later but not immediate text, or (c) the same form was produced by the same learner in two (or more) different texts.

A summary of the results of revision with and without WCF on the accurate production of present subjunctive errors is shown in Table 4.29 below.

<table>
<thead>
<tr>
<th>Learner</th>
<th>WCF Group</th>
<th>No WCF Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L2 (n=7)</td>
<td>HL (n=1)</td>
</tr>
<tr>
<td>Revision + Immediate Production</td>
<td>- (N=0)</td>
<td>- (N=0)</td>
</tr>
<tr>
<td>Revision + Later Production</td>
<td>0% (N=8)</td>
<td>0% (N=1)</td>
</tr>
<tr>
<td>Revision + Production in two different texts</td>
<td>- (N=0)</td>
<td>- (N=0)</td>
</tr>
<tr>
<td>No revision + Immediate Production</td>
<td>0% (N=2)</td>
<td>0% (N=1)</td>
</tr>
<tr>
<td>No revision + Later Production</td>
<td>0% (N=2)</td>
<td>- (N=0)</td>
</tr>
<tr>
<td>No Revision + Production in two different texts</td>
<td>- (N=0)</td>
<td>- (N=0)</td>
</tr>
</tbody>
</table>

**Notes.** % = Percentage of correct use of present subjunctive; N = number of present subjunctive forms; Revision = a present subjunctive form was corrected during the revision session; No Revision = the present subjunctive form was not provided during the revision session; Immediate Production = the same present subjunctive form was used after the revision stage; Production in two different texts = the same present subjunctive form was used in two different texts; Later Production = the present subjunctive form was used in a subsequent, but not immediate, text.

The results of the L2 learners in the WCF group revealed no instances of traceable forms of present subjunctive (N=0) that were revised and used immediately in a new text. Additionally, no traceable forms (N=0) of present subjunctive error that were correctly revised and produced in
two different texts were found in the data of the L2 learners in the WCF group, as shown in Table 4.29.

The results indicate that the present subjunctive errors that were revised and produced later in new pieces of writing (N=8) exhibited an accuracy rate of 0%. With respect to the present subjunctive forms that were not revised by the L2 learners, the results indicate that there was no correct use of when the same form (N=2) was used immediately (0%) and neither when the same present subjunctive form (N=2) was produced in a later text (0%). Furthermore, the results in Table 4.29 show only one traceable form in the HL learners in the WCF group (N=1), which consisted of an unrevised present subjunctive form (N=1) that was used incorrectly in immediate production (0%).

Figure 4.38
Effects of revision with and without WCF on the accurate production of present subjunctive forms (L2 and HL learners)

With respect to the L2 learners in the No-WCF group, the results in Table 4.29 and Figure 4.28 show no correct use (0%) in either those forms that were previously revised and then used in immediate production (N=1) and with revised forms that were produced in a later text
Additionally, the results indicate that when the present subjunctive forms were not revised, the same form was always produced incorrectly by the L2 learners in either immediate (N=5) and later texts (N=9) produced by the L2 learners in later, but not immediate texts. Similarly, the results in Figure 4.38 reveal that when the HL learners did not revise their present subjunctive errors, the correct form was never produced (0%) in later texts (N=6).

4.4. Research question 3a: Complexity and Fluency Results

The third research question was to determine whether there were any differences between the WCF and No-WCF conditions in other areas of writing development as a result of the treatment. As was described in Chapter 3, the components of writing development that were analyzed included two measures of complexity (i.e., verbal density and lexical richness) and one measure of fluency (average number of words per minute). Research question 3 asked whether differences existed before and after the intervention between the two group conditions in terms of (a) complexity and (b) fluency development.

4.4.1 Research question 3a: Complexity Results

The results revealed similar complexity scores between the WCF and No-WCF groups before and after the treatment, as shown in Table 4.30 and Table 4.31 below.

<table>
<thead>
<tr>
<th>Table 4.30</th>
<th>Effects of intervention on Verbal Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal Density</td>
<td>WCF</td>
</tr>
<tr>
<td>Pre-intervention</td>
<td>M= 2.71</td>
</tr>
<tr>
<td></td>
<td>(SD= 0.49)</td>
</tr>
<tr>
<td>Post-intervention</td>
<td>M= 2.36</td>
</tr>
<tr>
<td></td>
<td>(SD=0.57)</td>
</tr>
</tbody>
</table>

Notes. Verbal Density = Total number of finite and non-finite verbs divided by the total number of t-units; WCF = revision with corrective feedback; No WCF =
The results in Table 4.30 indicate similar mean scores in verbal density between the WCF group (M=2.71; SD= 0.49) and the No-WCF (M=2.73; SD= 0.8) before the treatment. The results also revealed that after the treatment period both the WCF group (M=2.36; SD=0.57) and the No-WCF group (M=2.39; SD= 0.48) experienced a decrease in verbal density, as shown in Figure 4.39.

A two-way repeated measures ANOVA was conducted to assess whether the differences before and after the intervention in terms of verbal density were statistically significant between the two conditions. The results revealed no significant differences on the measure of verbal density between the WCF and No-WCF groups, Wilk’s Lambda= 1.0, F(1, 34) = 0.001, p= 0.98 partial η² = 0.000.

With respect to second measure of complexity, the results revealed that the WCF and No-WCF groups increased their lexical richness scores at the end of the treatment, as shown in Table 4.31 below.
Table 4.31
Effects of intervention on Lexical Richness

<table>
<thead>
<tr>
<th></th>
<th>WCF</th>
<th>No WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lexical Richness</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-intervention</td>
<td>M=6.12</td>
<td>M=6.21</td>
</tr>
<tr>
<td></td>
<td>(SD=0.78)</td>
<td>(SD=0.8)</td>
</tr>
<tr>
<td>Post-intervention</td>
<td>M=6.14</td>
<td>M=6.35</td>
</tr>
<tr>
<td></td>
<td>(SD=0.67)</td>
<td>(SD=0.52)</td>
</tr>
</tbody>
</table>

Notes. Lexical Richness = Number of types divided by the square root of the number of tokens; WCF = revision with corrective feedback; No WCF = revision without corrective feedback; pre-intervention = session 1; post-intervention = session 11.

The lexical richness scores were similar in the WCF group (M=6.12; SD=0.78) and the No-WCF groups (M=6.21) before the intervention. However, the No-WCF group experienced higher gains in lexical richness after the intervention (M=6.35; SD=0.52) as compared to the WCF group (M=6.14; SD=0.67), as shown in Figure 4.40.

Figure 4.40
Interaction between group and time (Lexical Richness)

A two-way repeated measures ANOVA was calculated to determine whether the differences in terms of lexical richness between the WCF and No-WCF groups were statistically
significant before and after the intervention. The results revealed that the two groups did not
differ significantly across the two time periods in the measure of lexical richness, Wilk’s
Lambda=0.99, \( F(1, 34) = 0.16, p= 0.69 \) partial \( \eta^2 = 0.005. \)

4.4.2 Research question 3b: Fluency Results

The fluency results displayed in Table 4.32 show differences between the WCF and No-
WCF groups before and after the intervention.

Table 4.32

<table>
<thead>
<tr>
<th>Fluency</th>
<th>WCF</th>
<th>No WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-intervention</td>
<td>M= 15.8</td>
<td>M= 17.06</td>
</tr>
<tr>
<td></td>
<td>(SD= 4.48)</td>
<td>(SD=4.62)</td>
</tr>
<tr>
<td>Post-intervention</td>
<td>M= 17.62</td>
<td>M= 16.1</td>
</tr>
<tr>
<td></td>
<td>(SD=4.31)</td>
<td>(SD = 4.15)</td>
</tr>
</tbody>
</table>

Notes. Fluency = Number of types divided by the square root of the number of
tokens; WCF = revision with corrective feedback; No WCF = revision without
corrective feedback; pre-intervention = session 1; post-intervention = session 11

The results indicate that the average number of words per minute was lower in the WCF
group (M=15.8; SD= 4.48) as compared to the No-WCF group before the treatment. A pre-post
intervention comparison revealed that the WCF group increased their fluency over time
(M=17.62; SD= 4.31), while the No-WCF group experienced a decrease (M=17.06; SD=4.15),
as shown in Figure 4.41 below.
A two-way repeated measures ANOVA was computed to determine whether the variation between the WCF and No-WCF was statistically significant. The results of the observed F value indicate that the differences between the two groups across the two time periods in the measure of fluency were not significant, Wilk’s Lambda=0.90, $F(1, 34) = 3.6$, $p= 0.07$ partial $\eta^2 = 0.096$. 

Figure 4.41
*Interaction between group and time (Fluency)*

![Graph showing interaction between group and time for fluency](image_url)
CHAPTER 5: DISCUSSION AND CONCLUSION

5.1. Introduction

This study contributes to the field of SLA by providing empirical evidence to the general question of whether comprehensive written corrective feedback facilitates language learning by comparing the extent to which both error revision with and without WCF lead to knowledge retention and acquisition of linguistic forms over time. This chapter begins with a discussion of the results from the first research question which centered on the effects of error revision with and without WCF on the learners’ ability to self-correct their own errors and to correctly produce specific L2 forms in new pieces of writing, including (a) canonical gender marking, (b) non-canonical gender marking, (c) definite articles in obligatory contexts and (d) the present subjunctive. Next, the findings of the second research question, namely, the comparison of the effects that the two conditions had on both L2 and HL learners, will be discussed attending to the similarities and differences found in the current literature between these two populations. The third question, that is, the impact of the WCF treatment in the areas of fluency and complexity, will be discussed drawing from the available research on language development and models of attention. This discussion section will conclude with the limitations of the current study and present directions for future research on the effects of grammar correction and L2 written production.

5.2. WCF and No-WCF group error revision and accuracy development

In order to answer the first research question, the production, error revision and accurate production of each of the four linguistic forms, including (a) canonical gender marking, (b) non-
canonical gender marking, (c) definite articles in obligatory contexts and (d) present subjunctive, were compared between the WCF and No-WCF group.

5.2.1. Canonical Gender Marking

Before discussing the results regarding the differences between the WCF and No-WCF groups in terms of revision and accurate production of canonical gender marking from an SLA standpoint, the findings regarding the production and distribution of canonical gender marking need to be considered. The production results showed that the mean production and frequency of canonical gender assignment was much higher than that of canonical gender agreement (2) in the data of the two groups.

(1) \textit{El mundo necesita más} \\
‘The [masc] world [masc] needs more’

(2) \textit{Una vida buena} \\
‘A [fem] better life [fem]’

In example (1) the canonical gender ending noun \textit{mundo} ‘world’ (masc.) was assigned the correct gender through the definite article \textit{el} ‘the’ (masc.), whereas in example (2) the adjective \textit{buena} ‘good’ (fem.) agreed in gender with the canonical ending noun \textit{vida} ‘life’ (fem.). Differences in the distribution of canonical gender marking were expected considering that noun-adjective concord (i.e., gender agreement) is far less common in the production of Spanish than determiner-noun concord (i.e., gender assignment), because determiners are typically a necessary component in the Spanish noun phrase (NP). The kinds of determiners found in the written production of the learners included definite and indefinite articles, as in examples (1) and (2) above, demonstratives (3), possessive determiners (4), and quantifiers (5), as in the examples below.
It should be also noted that the large differences found in the gender marking distribution could be also attributed to the use of argumentative prompts, as in (6), which, as opposed to narration or description, might have contributed to the elicitation of fewer cases of noun-adjective production.

(6) ¿Crees que hay igualdad entre los hombres y las mujeres? (Session 3)
‘Do you think there is equality between men and women?’

Comparison of Canonical Gender Marking Error Production

Regarding the overall production of gender marking errors, the results showed that the frequency of assignment errors, as in (6), was lower than that of gender agreement errors (7) in both groups, regardless of the treatment condition.

(7) En este vida necesitamos dar
‘In this [masc]* life [fem] we need to give’

(8) Hay personas que son muy ricos
‘There are people [fem] that are very rich [masc]*’

In example (7) the canonical gender ending noun vida ‘life (fem.)’ was assigned the correct gender through the demonstrative article este ‘this (masc.), while in example (8) the adjective ricos ‘rich (masc.)’ did not agree in gender with the canonical ending noun personas
‘people’ (fem.). These results are in line with the findings of previous experimental studies (e.g., Bruhn de Garavito & White, 2002; Finnemann, 1992; Fernández-García, 1999; Martinez-Gibson, 2011) that reported that L2 learners tend to be less accurate with gender concord with adjectives than with determiners.

Comparison of Canonical Gender Marking Error Revision

With respect to the learners’ ability to self-correct their canonical gender marking errors, the total and mean number of gender assignment (9) and gender agreement (10) revisions made by the participants in the WCF and No-WCF groups throughout all the nine revision sessions were compared.

(9) \( \text{algo como } \text{un (una) casa} \)

‘something like a \([\text{masc}]^{*} \text{(a) [fem] casa [fem]}\)’

(10) \( \text{Es algo que debe ser } \text{publica (publico)} \)

‘It is something \([\text{masc}]\) that should be \text{public [fem]}^{*} \text{(public) [masc]}\’

In example (9) a canonical gender assignment error \(*\text{un casa} \ ‘a (masc.) house (fem.)’\) was successfully revised by adding the correct indefinite article \(\text{una} \ ‘a’ \ (\text{fem.}) \) next to the original not target-like form that was marked in boldface. In example (10) the adjective \(\text{publica} \ ‘\text{public’ (fem.)} \) did not agree with the canonical gender ending noun \(\text{algo} \ ‘\text{something’ and was marked in boldface. During the revision session, the correct gender form of the adjective \(\text{publico} \ ‘\text{public’ (masc.)} \) was added next to the original wrong form.

It was hypothesized, based on previous research that has investigated students’ ability to revise their own errors by means of indirect WCF feedback (e.g., Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001), that the participants in the feedback group would provide a larger number of accurate self-revisions of their errors. As expected, the learners
in the WCF group were able to provide significantly more correct revisions in both canonical gender assignment and canonical gender agreement errors as compared to their counterparts that did not receive form feedback. These findings offer further evidence for the efficacy of indirect CF as an instrument to guide and enable students to self-correct their errors with canonical gender ending nouns during revision.

**Comparison of Canonical Gender Marking Accurate Production and Development**

While the number and percentage of successful revisions provide indication that the learners have noticed the errors in their own output and that they are able to provide the target-like version, as some authors (Bitchener, 2012; Bruton, 2009a; Polio et al., 1998; Truscott & Hsu, 2008) have explained, the revision of the errors on a text does not entail evidence of learning or second language acquisition. Consequently, in order to provide evidence on the effects of comprehensive WCF on the acquisition of canonical gender marking and how it compares to text revision without form feedback, accuracy scores for both canonical gender assignment and agreement were calculated for all the learners in the feedback and no feedback conditions.

The mean accuracy results and the developmental trends for each gender marking type revealed some similarities and differences between the two groups that are worth discussing. First, the WCF and No-WCF groups were highly accurate in the production of canonical gender assignment, including many participants that consistently scored at ceiling. Although the purpose of this study is not to test whether adult learners whose L1 does not have gender (e.g. English speakers learning Spanish) can acquire the gender feature nor whether the acquisition of gender is subject to maturational constraints, the fact that some of the learners in this study performed with native-like accuracy in timed written production supports the view that the canonical gender
assignment feature might be acquirable in L2 acquisition, and this adds to the findings of previous studies that employed offline written production and comprehension tasks (Alarcon, 2011; Gruter et al., 2012; Montrul et al., 2008; White et al., 2004). Additionally, as compared to gender assignment, the lower mean accuracy scores in gender agreement provide indication that learners may undergo more difficulties in correctly producing this feature, whether due to maturational constraints (Hawkins & Franceschina, 2004) or insufficient linguistic experience or exposure (White et al., 2004).

With respect to the efficacy of the WCF treatment to foster the accuracy development of canonical gender marking, it was found that, while the group that did not receive corrective feedback experienced a small decrease in terms of accurate production of canonical gender assignment, the accuracy trend in the WCF group remained stable; however, in terms of canonical gender agreement, the opposite pattern was found as the accuracy performance of the treatment group slightly decreased over time, whereas the accurate production in the no-feedback group remained steady. There are two potential explanations that may account for these two opposing patterns in canonical gender marking. First, considering that the amount of revision of gender assignment and agreement errors was comparable in the treatment group, it seems that the WCF was only effective in helping develop the accuracy of canonical gender assignment, as shown by the trajectories followed by each group. It is possible that learners might require further input and negative evidence in order for the WCF to have an effect on the acquisition of canonical gender agreement, given that, as research on the acquisition of Spanish gender by English speakers has shown (e.g., Bruhn de Garavito & White, 2003; Fernandez-Garcia, 1999), L2 learners of Spanish are less accurate on gender concord with adjectives as this feature is acquired later than determiner-noun assignment.
It is important to note, however, that, while the two groups followed different directions in terms of gender marking development, the mean differences in accuracy were indeed rather marginal, making it difficult to actually discern whether such variation, especially in canonical gender assignment, are due to the WCF treatment. In this particular case, a closer look at intra-individual performance and variability, as Larsen-Freeman (2006) argued, can offer better insight into the underlying developmental process and provide further evidence on the efficacy of the treatment, which group data that is averaged may obscure. The individual patterns found in the WCF group showed that, while it is possible that some learners might have benefited from the CF, there is evidence that the WCF treatment did not have a differential effect on canonical gender assignment and, especially, gender agreement given the large number of participants that exhibited a decrease in accuracy over time. The limitation of the treatment is further attested by the results of the individual patterns in the No-WCF group which showed that a large number of participants were able to increase their gender marking accuracy over time without the need for corrective feedback. A possible reason why some learners were able to improve their accuracy without negative evidence may be related to the role of output in the acquisition of partially-acquired forms. For example, Swain (1995) argued that output could be beneficial to stimulate learners “to move from the semantic, open-ended nondeterministic, strategic processing prevalent in comprehension to the complete grammatical processing needed for accurate production” (p.128). Likewise, within the writing-to-learn perspective, L2 writing is seen as an instrument that can facilitate L2 development and some authors have suggested (Manchón, 2011; Williams, 2012) that practice alone might be sufficient to help L2 learners internalize, restructure and consolidate L2 knowledge or forms that are already part of their developing system (Housen & Pierrard, 2005). In this sense, the improved accuracy of some of the learners in the No-WCF
group, as shown by their developmental trends, suggests that written production could have contributed to the third process of L2 development, that is, knowledge consolidation, by facilitating the strengthening of their canonical gender marking knowledge through repeated retrieval and deeper processing. However, a follow-up of this study with data collected but not analyzed that include the recording of the learners’ writing process in real time or the use of other introspective measures such as think-alouds or retrospective questionnaire data would be very helpful to provide further evidence of the extent to which written production alone may consolidate L2 knowledge.

5.2.2. Non-Canonical Gender Marking

The distribution of non-canonical gender marking in the free-written production of the participants in the WCF and No-WCF groups revealed that the occurrence of non-canonical gender assignment (1) was significantly higher than non-canonical gender agreement, as in (2).

1. \textit{Es un problema grande}  
   ‘It is a [masc] big problem [masc]’

2. \textit{Puede crear oportunidades económicas}  
   ‘It can create economic [fem] opportunities [fem]’

These results on non-canonical gender marking production are similar to those found earlier on canonical gender marking, as the occurrence of determiner-noun concord was considerably more frequent than that of noun-adjective. As was argued in the previous section, the differences in the frequency and distribution are likely to be due to the fact that NPs in Spanish typically include both a determiner and a noun, which must agree in gender and number. In addition, the argumentative topics used to elicit the written production may have also influenced the distribution of canonical gender marking as noun-adjective concord, which
constitutes gender agreement, typically appear less frequently in argumentation, as compared to other genres such as description.

**Comparison of Non-Canonical Gender Marking Error Production and Revision**

The overall production of gender marking errors is consistent with what was found in canonical gender marking, that is, assignment errors (3) were less frequent than agreement errors

(3)  *Los accion*enes
    ‘The [masc]* actions [fem]’

(4)  *País*es extranjer*as*
    ‘Foreign [fem]* countries [masc]’

Likewise, the distribution of non-canonical gender marking errors was comparable to what was found in regular gender marking, that is, a higher proportion of gender agreement errors in both the WCF and No-WCF groups. While the production of non-canonical gender agreement only comprised between 10% and 12% of all of the gender marking production; agreement errors accounted for 30% of all of the gender marking error in the WCF group, and 22% in the No-WCF group.

Furthermore, a comparison in terms of the total number of errors with canonical and non-canonical gender ending nouns revealed that non-canonical gender marking errors were more frequent than canonical gender marking errors in the written production of the two groups, including gender assignment and agreement. These findings from free written production are in line with previous experimental studies (Alarcón, 2011; Montrul et al., 2008; Montrul et al., 2014) that found that learners of Spanish were more inaccurate with non-canonical than with canonical gender nouns in written comprehension, written recognition, oral production and grammaticality judgment tasks. As mentioned earlier, in Spanish, canonical gender is construed
by adding the inflectional morpheme ‘o’ to masculine nouns as in niño (masc) and ‘a’ to feminine nouns niña (fem.). However, there are also many exceptions or ‘irregularities’ in the system and some masculine and feminine nouns can end in the vowel ‘a’ or ‘o’, and also in the vowel ‘e’, or with a consonant. According to Montrul et al. (2013), differences in gender marking accuracy between canonical and non-canonical gender ending nouns are primarily due to the morphological ambiguity of non-canonical nouns, which creates a problem for adult learners.

In terms of the learners’ ability to self-correct their non-canonical gender marking errors, the average and total number of gender assignment (5) and gender agreement (6) errors that were revised by the WCF and No-WCF groups were compared.

5. *En otros (otras) partes del mundo*

‘In other [masc]* (other) [fem] parts [fem] of the world’

6. *Una acción sencillo (sencilla)*

‘An [fem] easy [masc]* (easy) [fem] action [fem]’

As expected, the group that received indirect WCF feedback was able to provide significantly more revisions than the group that did not receive any error feedback. More specifically, the results showed that the learners in the WCF group correctly revised 80% of all of their assignment errors and 81% of their gender agreement errors, while their counterparts in the no feedback condition were only able to revise 26% of their gender assignment errors and 28% of their agreement errors throughout all nine revision sessions. These findings add to previous studies on error revision (e.g., Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001) that have indicated the efficacy of indirect WCF as a tool to help learners self-correct their errors of a previous text, and also showed that, in absence of corrective
feedback or negative evidence, intermediate-advanced learners of Spanish have difficulties noticing and self-correcting their non-canonical gender errors.

**Comparison of Non-Canonical Gender Marking Accurate Production**

As mentioned earlier, non-canonical gender, unlike canonical gender marking which follows formal rules (i.e., ending ‘-o’ for masculine and ‘-a’ for feminine), is subject to various irregularities (-e, consonant, opposite vowel) and, therefore, is not transparent to L2 learners. For this reason, it was argued in previous chapters that since there are many different non-canonical nouns, incorporating all of the non-canonical gender marking errors into one measure would fail to provide an accurate picture of the treatment and obscure cases in which the learning of a particular non-canonical form may (not) have occurred. Hence, it appears self-evident that the best or at least the most precise way to provide evidence that a WCF treatment helps increase the accurate production of non-canonical gender marking (or any non rule-based and/or complex forms) is by looking at whether forms that were previously revised (or not) are produced correctly in new pieces of writing.

For the purpose of this study, the production of non-canonical gender ending nouns that were traceable, meaning that they were used in more than one text, were classified according to whether the same revised form was (a) immediately produced by the same learner in the following text, (b) produced by the same learner in a later but not immediate text, or (c) produced in two (or more) different texts by the same learner.

As predicted, the highest rate of correct use of non-canonical gender marking forms (83.3%) occurred with participants in the WCF group who, after successfully revising an error, used the same forms in immediate production, as in Example (1).
Example 1

(WCF)S9  El dinero no para las (los) problemas  
(Session 6)

‘Money does not stop the [fem]* (the) [masc]
problems [masc]’

(WCF)S9  Yo creo que hay un problema  
(Session 7)

‘I think there is a [masc] problem [masc]’

In example (1), the non-canonical noun problemas ‘problems’ (masc.) was assigned the wrong gender through the definite article las ‘the’ (fem.), and, consequently, the not target-like form (la ‘the’–fem.) was marked in boldface. In the next session (Session 7), the learner noticed the error and self-corrected it by adding the correct form, los ‘the’ (masc.) in parenthesis next to the original erroneous form in boldface. After the revision session, an instance of correct gender assignment with the noun problema ‘problem’ (masc.) and the indefinite article un ‘a’ (masc.) was produced in the new piece of writing produced immediately thereafter. The fact that most learners were able to use the correct non-canonical gender assignment form in immediate production is not surprising considering that the information obtained from the WCF and the processing of their errors was very recent and presumably still available in their working memory when they were writing their new text.

With respect to the learners in the No-WCF group, the results showed that non-canonical gender assignment forms that were not revised but subsequently produced in immediate production, as in Example 2, were used correctly sixty percent of the time.

Example 2

(No-WCF)S4  Un estereotipo en la programa  
(Session 2)

‘A stereotype in the [fem]* program [masc]’
The extract from Example 2 shows that an incorrect definite article, *la* ‘the’ (fem.), was assigned to the non-canonical noun *programa* ‘program’ (masc.). Although this non-canonical gender assignment error was not revised during the revision session, the correct non-gender assignment form *un programa* ‘a program’ (fem.) was correctly used in the following piece of writing (Session 3). There are some possible explanations for why some learners, after producing a non-canonical gender error, were able to use the same form correctly in a new piece of writing. First, as mentioned earlier, it is possible that the learners had the explicit grammatical knowledge but failed to retrieve the correct information at one particular time, which occurs when there is not full control over a specific linguistic form. It is also possible that, as part of the process of linguistic restructuring, both the target and not target-like forms are still part of the learner’s developing grammar. That is, their choice of one form or the other might still be variable and probabilistic. With respect to why most learners in the No-WCF group, in spite of being able to use the correct form in subsequent production, did not self-correct their errors during the revision sessions, again it is difficult to determine whether the learners overlooked or decided to ignore a certain error and focus their attention on some other component of their writing when they were making revisions. However, the fact that the learners still had competing forms (i.e., grammatical and ungrammatical) could help explain why most of them were not able to detect their non-canonical gender errors and make any revisions.

Hence, in order to test the long-lasting effects of the WCF on non-canonical gender assignment forms, the rate of correct use of those forms that were successfully revised and used,
not immediately, but in later production, as in Example 3, can offer further evidence of the effects of the treatment.

Example 3

(WCF)S18  *Representa un (una) acción*  (Session 2)

‘It represents an [*masc]* (an) [*fem]* action [*fem]*’

(WCF)S18  *Las malas acciones que alguien hace*  (Session 11)

‘The [*fem]* bad actions [*fem]* that someone does’

The results showed that the participants in the WCF group were able to produce the correct non-canonical form in a later text over fifty percent of the time (57.1%). It important to note that while these results may provide some indication of the effects of the feedback, still it is not possible to determine whether some of the errors were the result of the learner not knowing the correct gender of the noun or the learners’ failure to retrieve the correct gender during production.

With respect to the effect of revision on the participants in the No-WCF group, the results revealed a higher rate (75%) of correct use of forms that were revised and incorporated in later production as compared to the WCF group (57.1%). These findings are not surprising because the fact that the learners in the no-feedback condition were able to notice their errors and provide the correct forms during the revision session, as in Example 4, implies that they already had knowledge of the correct gender of the non-canonical noun.

Example 4

(No-WCF) S1  *Pienso que las (los) problemas del mundo*  (Session 9)

‘I think that the [*fem]* (the) [*masc]* problems [*masc]* of the world’
‘Yes, there is a [masc] problema [masc] with’

The excerpt in Example 4 shows that an error made in a previous text (Session 9) was successfully revised without any provision of feedback, and the same non-canonical form was correctly used in later production (Session 11). The ability to notice an error, self-revise it without any external feedback or negative evidence, and produce the same form correctly later in a new piece of writing could be seen as an indication that the error produced at an earlier stage was rather a performance problem, that is, a failure to retrieve the correct form during production, than one of competence as there is evidence that the learner may have the formal, or, in terms of skill acquisition theories (e.g., DeKeyser, 1997; Ullman, 2001), the declarative knowledge but have difficulty accessing it in spontaneous production.

Example 5

(No-WCF)S1  Con otras paí ses y gente en el mundo  

‘With other [fem]* countries [masc] and people in the world’

Example 5 shows that the non-canonical gender assignment error *otras países (other-fem. países-masc.), which was not revised, was repeated in a later production (Session 8).
These results were expected considering that if learners are not given any indication that their production is not target-like, it is then more likely that they will produce the same error in later production, as they might not be able to notice the gap between their IL and target language. In this regard, WCF is hypothesized to not only provide opportunities for learners to focus their attention on their errors, but also help make cognitive comparisons and noticing that can potentially facilitate language acquisition (Ellis, 1994; Schmidt & Frota, 1986). Hence, if when assessing the effects of the treatment on accuracy development we consider revised and non-revised forms as indication that an error has been noticed and to some extent processed, the results could be interpreted as indication that the revision of non-canonical errors was more effective as it led to more accurate use of the same form in later production (57%) as compared to those forms that were not previously corrected (23.1%).

However, in addition to forms that were used in later production, it could be argued that the correct use of the same form in multiple pieces of writing, as in Example (6), could provide more solid evidence on the extent to which the effect of the feedback is maintained over time.

**Example 6**

(WCF)S17 *Las armas de fuego son una (un) problema […]* (Session 7)

*es una (un) problema que puede […] No podemos arreglar la (el) problema*

‘Fire arms are a [fem]* (a) [masc] problem [masc] it is a [fem]*

(a) [masc] problem [masc] that can… We cannot fix the [fem]*

(the) [masc] problem [masc]’

(WCF)S17 *Los problemas hoy con los jóvenes* (Session 8)

‘The [masc] problem [masc] today with the youth’
(WCF)S17  *Especialmente el problema de emigrantes ilegales*  (Session 11)

‘Especially the [masc] problem [masc] with illegal immigrants’

While, Example 4 offers an instance of the correct use of gender assignment with the non-canonical noun *problema* ‘problem’ (masc.) in two different pieces of writing (Session 8 and Session 11) after being previously revised, the results indicated a very low rate of correct use (14.3%) when the same non-canonical gender assignment form was produced in a third text, as the excerpts in Example 7 show.

**Example 7**

(WCF)S3  *Hay una (un) problema con los manejadores*  (Session 8)

‘There is a [*fem]* (a) problem [masc] with drivers

(WCF)S3  *Es un gran problema y tiene un gran impacto*  (Session 9)

‘It is a [masc] great problem [masc] and it has a great impact’

(WCF)S3  *Pero los problemas pueden resolver*  (Session 10)

‘But the [*fem]* problems [masc] can solve’

A possible explanation for this failure to provide the correct gender assignment over time may relate to the complex process of acquiring L2 forms. For example, in terms of the skill acquisition models for learning (e.g., Anderson, 1983, 1987; McLaughlin, 1987, 1990) learning is seen as a cognitive process that requires constant restructuring as learners tend to simplify or unify their mental representations until they gain control over them. In this view, it is then not realistic to assume that once learners receive corrective feedback, they will always instantly be able to gain full control of a grammatical feature. Instead, during the process of restructuring or
IL development that may lead to the L2 acquisition, it is not uncommon that both the target and
not target-like forms coexist, as appeared to be the case in Example 8.

Example 8

(WCF)S7  Y para la mayoría no es una (un) problema  (Session 7)
           ‘And for the majority it is not a [fem]* (a) [masc]
           problem [masc ’]

(WCF)S7  ‘Hay demasiados problemas […]
           Este problema ya es un problema con las personas […]
           podría un problema para los estudiantes’
           ‘There are too many [masc] problems [masc] This [masc]
           problem [masc] is already a [masc] problema [masc] with people’

(WCF)S7  Los economicos son una problema […]  (Session 9)
           son un problema también
           ‘The economics are a [fem]* problema [masc] They are a [masc]
           problem [masc]’

In Example 8, a participant in the WCF group first revised a gender assignment error with
the non-canonical ending noun problema ‘problem’ (masc.) by adding the correct indefinite
article el ‘the’ (fem.), next to the erroneous form marked in boldface. After the revision session,
in a succeeding new text (Session 8) the learner was able to assign the correct gender to the noun
‘problema’ (fem.) using different types of determiners, including demasiados ‘too many’
(masc.), un ‘a’ (masc.) and este ‘this’ (masc.). However, in the next session (Session 9), the
learner produced both the non-target *una problema (‘a-fem. problem-masc.’) and the target-like
gender assignment forms un problema (‘a-fem. problem-masc.’) in the same text, which shows
that the gender feature for the non-canonical noun problema ‘problem’ had not yet been
acquired. Some researchers (DeKeyser, 1993; Ellis, 2005) have argued that retrieving and using explicit knowledge can be automatized through prolonged practice (i.e., production), which can in turn facilitate the internalization and consolidation stages identified by interaction theories. However, an important question that needs to be considered is the extent to which learners, after they receive feedback, are able to incorporate it in their production. A corpus analysis of the feedback and production in the WCF group indicated that although the learners revised a total of 72 non-canonical gender marking forms, only 28% of those forms were traceable, meaning that they were used again by the same learner in new pieces of writing. In other words, it is possible that the majority of the feedback (72%) that the learners received for non-canonical gender marking and the errors that they revised, as in Example 9, were ineffective in terms of acquisition, not because of the lack of opportunities for production, but because the same form, for one reason or another, was not produced again, and as such the conditions necessary to help develop their linguistic knowledge (i.e., prolonged practice, feedback, etc.) were not met. On the other hand, it is also possible that the learners in the WCF group did learn and retain the knowledge from at least some of those corrections; however, there is a problem of ‘absence of evidence’ because of those forms that were not used again during production. A possible solution for this problem could be the use of tailor-made posttests to test the retention of specific forms, but they would in turn detract from the authenticity of free-production tasks from a classroom composition setting.

5.2.3. Definite Articles in Obligatory Contexts

The purpose of this research question was to examine the extent to which WCF can help learners revise their article omission errors and whether error revision helped the learners become more accurate with the written production of definite articles in Spanish.
Before addressing these questions, it is important to briefly review some of the differences between English and Spanish in the structure of determiner phrases (DP) that contribute to the misuse of definite articles in the production of learners of Spanish whose first language is English.

According to Chierchia (1998) languages may differ in that the structure of their nominals may or may not require determiners. Languages that are [+arg], such as English, allow bare nouns to be arguments (i.e., a noun that can denote a kind) in the syntax (1).

(1) ‘Ignorance is bliss’

On the other hand, languages that are [-arg], such as Spanish, do not allow bare nouns in the syntax to be arguments, as in (2).

(2) *Ignorancia es un problema

Ignorance is a problem

‘Ignorance is a problem’

As mentioned in chapter 3, one of the problems for L1-English L2-Spanish learners is that, while English uses bare nouns to express genericity and DPs to express specificity, Spanish uses the article form to encode both the generic and specific meanings. As a result, learners of Spanish transfer the generic meaning to Spanish by omitting the definite article (Ionin & Montrul, 2010; Montrul & Ionin, 2012). Although there are some exceptions that are outside the scope of this study, there is general agreement that definite articles are required before generic nouns, including abstract nouns referring to general concepts (3), countable nouns that refer to all members of a class (4) and substances in general (5), which differ from English in that the generic reference can be expressed with bare plurals in NPs.

(3) Los estereotipos son parte del problema
‘Stereotypes are part of the problem’

(4)  *Los niños duermen más que los adultos*

‘Children sleep more than adults’

(5)  *El agua es muy importante*

‘Water is very important’

In terms of production, it was found that the overall number of generic nouns that required the use of definite articles (i.e., definite articles in obligatory contexts), in addition to the average number of cases per text, were similar between the WCF and No-WCF groups, which in turn favors the comparison between the two groups.

*Comparison of Definite Article Omission Error Production*

A comparison between the WCF and No-WCF group revealed significant differences between the two conditions in terms of total errors of definite article omission. The participants in the no-feedback condition produced approximately double the amount of errors (N=187) as their counterparts (N=96) who did not receive feedback across all sessions.

*Comparison of Definite Article Omission Error Revision*

Given the differences in the total number of definite article omission errors between the two conditions, the proportion of errors that each group was able to revise with WCF and without WCF was calculated in order to measure the effects of the treatment on error revision, as in Example 1.

**Example 1**

(WCF) S9  *(article omitted)*  *Los Hombres generalmente tienen*  (Session 4)

◎(The) men generally have
Men generally have

(No-WCF) S4 (Los) *Hombres reciben más dinero*  
Ø(The) men receive more money  

‘Men receive more money’

In Example 1, a participant in the WCF group received WCF for an error of definite article omission, as indicated by the information in boldface next to the generic noun *hombres* ‘men’, and added the missing definite article *los* ‘the’ (masc. pl.) during the revision session. The participant in the No-WCF group (S4) was able to provide the correct definite article *los* ‘the’ (masc. pl.) next to the generic noun *hombres* ‘men’ without the help of WCF.

Similar to the revision of gender marking errors, it was hypothesized that the participants in the WCF group would provide a larger number of correct self-revisions of their article omission errors as compared to their counterparts in the WCF group. The results showed that the proportion of errors that the participants in the feedback group revised successfully was much higher (95%) compared to those in the No-WCF group (6%). These findings provide evidence that the indirect CF was effective in helping learners self-correct their errors of definite article omission, while, at the same time, they also point that, without any feedback help, the learners were rarely able to notice and correct these errors in their own written output. With respect to the efficacy of the treatment for error revision, it should be noted that although the type of correction for definite article omission was still indirect in nature, namely, the error was marked in boldface and the student needed to work out the correct target form, given that there was an indication of the error type (‘article omitted’) next to the error, it is possible that the learners benefited from that explicit information, which may have resulted in a higher rate of successful revisions as compared to other targets where explicit information was not provided. The decision to include
metalinguistic information about the error type was made based on information obtained in a pilot study conducted at an earlier stage that indicated that learners had difficulty identifying their article omission errors without explicit information.

**Comparison of Accurate Production of Definite Articles in Obligatory Contexts and Accuracy Development**

As argued earlier, in order to assess the efficacy of a feedback treatment on the accuracy development of specific linguistic forms, it is necessary to track the changes in accuracy over time and in multiple new pieces of writing (e.g., Bitchener, 2012; Bruton, 2009a). As a result, the accuracy scores for the use of definite articles with generic nouns in obligatory contexts were tracked over the course of four weeks constituting a total of eleven sessions.

The accuracy rate and developmental trends calculated for each group and all learners revealed some interesting information that should be discussed. In terms of the accurate use of definite articles in obligatory contexts, the results showed that the accuracy trend of the WCF group increased considerably during the data collection period, whereas the developmental trend in the no-feedback group remained stable over time. These group results seem to be congruent with what WCF advocates claim, that is, that frequent feedback and practice lead to higher accurate production, whereas the lack of feedback caused the control group not to notice their errors during revision, which hindered their accuracy development over time. While this could have been the case, the individual results showed a more complex picture of the effects of the two conditions on accuracy development, which needs to be considered.

With respect to the effect of the treatment on accuracy development, while it is clear that the indirect feedback assisted the learners in the revision of their errors, the low-medium effect size (d=0.34) of the treatment in the total number of omission errors suggests that it could have
actually not helped some learners reduce their article omission errors. This observation was further confirmed when the individual developmental trends for the production of definite articles in obligatory contexts were examined. A comparison between the two groups showed that, while the proportion of learners in the No-WCF group whose accuracy decreased over time was higher (38.9%) than in the WCF group (22.2%), the number of participants who increased their accuracy over time was exactly the same (44.4%). If we assess the efficacy of the treatment in terms of the number of the participants who became more accurate during the data collection period, then the differential effect of the treatment is not apparent.

Furthermore, the accuracy scores showed great variability per session, which ranged from 69.2% to 97.7% of correct use of definite articles in obligatory contexts in the WCF group and 64.2% to 89.8% in the No-WCF group. This variation and fluctuation in performance, as in Example 2, could be part of the interlanguage development instability, which as Larsen-Freeman (2006) noted, may precede a phase shift in the system during pedagogical interventions.

Example 2

(WCF)S1  
*Pienso que (omission of article) (la) tecnología hace*  
I think that *(the) technology makes*  
‘I think that technology makes’

(WCF)S1  
*Pienso que (omission of article) (el) dinero puede*  
I think that *(the) money can*  
‘I think that money can’

(WCF)S1  
*Pienso que la religión es*  
I think that the religion is  
‘I think that religion is’
In Example 2, two definite article omission errors with a similar structure (‘I think that + ☐ + generic noun’) were produced and revised correctly in two different texts (Session 5 and 6), and the definite article *la* ‘the’ was correctly placed before a generic noun *religión* ‘religion’ in a subsequent new piece of writing *Pienso que la religión* ‘I think that the religion’.

Another factor that needs to be borne in mind is the effect of priming when writers, or language users in general, answer a prompt or any input that is presented in the target language. In this case, the correct use of a form might not reflect the L2 learners’ grammatical knowledge, but would be the result of reproducing the same structure or forms seen in the input, as could be the case for Example 3 below, ¿*Crees que la sanidad debe ser pública en los Estados Unidos?* ‘Do you think that health care should be public in the US?’

**Example 3**

(NO-WCF) S9  *Sí, creo que la sanidad debe ser pública*

Yes, I think that the healthcare should be public

‘Yes, I think that healthcare should be public’

(WCF) S9  *Creo que la sanidad debe ser pública en los Estados Unidos*

I think that the healthcare should be public in the United States

‘I think that healthcare should be public in the United States’

(NO-WCF) S15  *Creo que la sanidad deba ser pública*

I think that the healthcare should be public

‘I think that healthcare should be public’

In Example 3, it is possible that the use of *creo que la sanidad* ‘I think health care’ was as part of a discourse strategy in which the learners decided to answer the question by reproducing the same structure or ‘language chunk’ observed in the input. If that was the case, then the
correct addition of the definite article *la* ‘the’ before the generic noun *sanidad* ‘healthcare’ may not have been part of their grammatical knowledge at that point, but rather the result of repeating the target form found in the input. While the question of whether this practice may facilitate learning is beyond the scope of this study, Lyster et al. (2013), in a study on oral production and recasts, argued for the potential learning effect of prompts in the target language, as the exposure to positive exemplars, in addition to practicing the target form, may help learners modify their output and aid in acquisition. In this sense, a comparison between the effects of negative (e.g., corrective feedback) and positive exemplars, such as the exposure and response to written prompts in the target language, on the acquisition/learning of L2 forms would be an interesting line of research that, to my knowledge, remains unexplored in writing research.

5.2.4. Present Subjunctive

This research question examined the extent to which comprehensive WCF, in addition to the explicit indication of the error type, helped learners self-correct their errors and whether their error revision led to subsequent accurate production of specific present subjunctive forms in new pieces of writing.

In Spanish, modality can be expressed grammatically in the verb, by means of mood morphology (i.e. indicative, subjunctive and imperative). In the case of subjunctive morphology (e.g., the present subjunctive), it appears most often embedded in complex sentences, including nominal (1), adjectival (2) and adverbial (3) clauses.

(1) *Es necesario que hagas eso*

‘It is necessary that you do [subj.]that’

(2) *Necesitamos trabajadores que puedan ayudar*

‘We need workers that can [subj.] help’
The structures that elicited the use of the present subjunctive were extracted and classified according to the type of subordinate clause indicated above (i.e., nominal, adjectival, and adverbial). Given that nominal clauses were far more frequent in the data, in order to provide a more detailed understanding of the type of production and errors made by the learners in this study, the use of the present subjunctive for this category was further classified according to the modality and meanings expressed in the main clause, which included (a) impersonal structures that express possibility, (b) verbs of doubt, (c) negated verbs or clauses, (d) impersonal structures that express opinion and (e) influence/volition.

a)  \textit{Es imposible que llegue hoy}

   ‘It is impossible that he arrives \textit{subj.} today’

b)  \textit{Dudo que pueda hacerlo}

   ‘I doubt that he can \textit{subj.} do it’

c)  \textit{No creemos que sea un problema}

   ‘We do not believe that it is \textit{subj.} a problem’

d)  \textit{Es importante que ayudemos}

   ‘It is important that we help \textit{subj.}’

e)  \textit{Quiero que me ayudes}

   ‘I want that you help \textit{subj.} me’

In terms of the production of the present subjunctive, the results showed that there were no major differences in the distribution of forms and structures that elicited the use of the present subjunctive across the WCF and No-WCF groups. Most of the cases that elicited the use of the
present subjunctive in both the WCF and No-WCF groups occurred in nominal clauses with verbs of denial/negation and impersonal structures that expressed opinion. A high occurrence of negated epistemic clauses, as in (d) above, was expected considering that its use would be the direct result of the learners taking a stance and expressing disagreement with some of the questions or statements included in the prompts. Example 1 below illustrates the case of a learner who expressed his opinion by means of negation.

Example 1

(WCF) S4 Y por eso no pienso que *son justos (Session 2)

‘And therefore I do not think [indic.] that they are fair’

Moreover, the high rate of impersonal structures that express opinion, as in (d) above, was not surprising as it can also be seen as the product of the learners choosing a position or stance when responding to the prompt. Additionally, the use of impersonal structures could be related to the use of perlocutionary acts in argumentative writing. As McCarthy and Carter (1994) noted, the writer’s personal voice is sometimes combined with impersonal advocacy, as part of a discursive style that intends to make arguments more objective, especially when trying to persuade or convince the reader, as in Example 2.

Example 2

(WCF) S16 Es integral que nosotros *ayudamos (Session 1)

‘It is integral that we help [indic.]’

These findings in terms of production suggest that the prompts employed in this study (see Appendix C) were effective in eliciting a comparable proportion and distribution of forms
and structures that required the use of the subjunctive, thus making the two WCF and No-WCF groups analogous.

Comparison of the Present Subjunctive Error Production

A comparison between the WCF and No-WCF groups revealed no differences in either the total number of errors across all the sessions or the overall percentage of accurate use of present subjunctive forms. The errors produced by the learners in the two groups primarily involved the use of the present indicative in contexts where the present subjunctive was required, as in Example 3.

Example 3

(WCF)S16  *Es posible que hay la misma*  (Session 9)

‘It is possible that there *is [indic.] the same’

(No-WCF)S15  *No importa que tienen fiestas*  (Session 6)

‘It does not matter that they *have [indic.] parties’

These types of errors, in which the present indicative is used in contexts where the subjunctive is required, were anticipated given that the subjunctive and indicative moods oppose each other and the learners, as part of the acquisition process, need to learn how to select them during production. As mentioned earlier in chapter 3, while explaining the complexity of the subjunctive is outside the scope of this study, it should be noted that its use is influenced by many factors, including lexical, syntactic, morphological and pragmatic features (see Collentine 2013 for a review), which learners need to acquire in order to understand and produce it accurately. Hence, if learners lack the linguistic knowledge for mood selection or fail to retrieve the correct form during production, errors such as the one in Example 3 may occur.

Considering the many variables involved in the acquisition of the present subjunctive, the
low accurate production of the present subjunctive is not surprising (33.5% correct use in the WCF group and 37% in the No-WCF group) because, as Montrul (2002, 2007) noted, the acquisition of the subjunctive is one of the most difficult aspects of the grammar of Spanish, which is also prone to fossilization in L2 learners.

With respect to the six categories that were found to elicit the production of the present subjunctive, the results provided some interesting information that is worth discussing. First, it was found that accuracy was the highest when the subjunctive was triggered by adverbial conjunctions in both the WCF (75%) and the No-WCF (85.7%) groups. While it is not possible to make any inferences from these results given the rather small sample of instances found in the data (N=15), the higher saliency of adverbial clauses could have contributed to the higher correct use, as compared to other structures (i.e., nominal and adjectival clauses) that may be less salient to the learners. For instance, when L2 learners receive instruction on the present subjunctive they learn that there are certain adverbial conjunctions (e.g., para que ‘so that’, antes de que, ‘before’ sin que ‘without’, etc.) that always introduce the use of the subjunctive in subordinate clauses, as in Example 4.

Example 4

(WCF)S17  para el alcohol antes de que tengas 21 años  (Session 15)
‘for alcohol before you are [subj.] 21 years old’

(No-WCF)S6  podemos cambiar para que no haya un problema  (Session 8)
‘we can change so there is [subj.] not a problem’

It is possible that the connection between certain types of adverbial conjunctions and the use of the subjunctive could have contributed by helping the learners process and retrieve the correct form during production as opposed to other contexts that semantically or pragmatically
select the use of the subjunctive such as volition, possibility or the contrast between realis-
irrealis and negation, which require more abstract representations and are therefore more
complex in terms of acquisition. In this regard, the results showed that the lowest level of overall
accuracy in both the WCF group (16.7%) and the No-WCF (12.5%) corresponded to negated
epistemic verbs, as in Example 5.

Example 5

(WCF) S15 no pienso el dinero da la felicidad (Session 6)
‘I do not think money *brings [indic.] happiness’

(No-WCF) S2 no creo que el dinero da la felicidad (Session 6)
‘I do not believe that money *brings [indic.] happiness’

Nonetheless, if we assume that the learners had already received instruction on the
subjunctive at some point, as they were sixth-semester students of Spanish, some possible
explanations for the low rate of correct use of the subjunctive with negated epistemic verbs could
be related to factors such as L1 transfer effect (Montrul, 2000, Montrul et al. 2008) or a
weakness in the syntactic-discourse/pragmatic interface as learners needed to interpret the
pragmatic relationship between the negator ‘no’ and the meaning of the verb in the main clause
(Iverson et al., 2008).

With respect to the overall accuracy results in the remaining categories, the No-WCF
group was more accurate with verbs of volition, impersonal structures that expressed possibility
and in adjectival clauses with an unknown antecedent, whereas the WCF group showed higher
accuracy with impersonal structures that expressed opinion. While no patterns in terms of
acquisition can be established due the written sample size, still the variation in the number of
errors and correct production suggest that not all the categories that elicited the use of the present
subjunctive were equally related. These findings indicate that the acquisition of the present subjunctive is not a ‘monolithic phenomenon’ (Collentine, 2003) and, as such, the efficacy of a pedagogical intervention, including corrective feedback, is likely to vary given the multiple forms and variables that may trigger the use of the subjunctive in addition to the current developmental stage of each learner.

With respect to the effects of the treatment in the total number and percentage of errors across all the sessions, the results did not show substantial differences between the two conditions. In fact, as mentioned earlier, the only category in which the WCF group (45%) was more accurate than the No-WCF group (23.1%) was with impersonal structures that expressed opinion. Hence, the findings clearly provide further indication that the WCF treatment did not make a difference in reducing the total number of subjunctive errors.

Another important question that needs to be considered regarding the treatment has to do with the ‘oversuppliance of the subjunctive’, that is, the use of the present subjunctive in contexts where the present indicative is required, as in Example 6.

Example 6

(WCF)S8  Creo que la sanidad sea muy importante  (Session 3)

‘I think that health care *is [subj.] very important’

(No-WCF)S12  Creo que muchas personas se olviden que  (Session 4)

‘I think that many people *forget [subj.] that’

Although it is difficult to clearly determine whether the extension in the use of the subjunctive, as in the previous example, is the result of lack of control or misunderstanding of the grammar rules, or simply part of interlanguage development, it is important to note that a potential adverse effect of focus-on-form interventions and explicit instruction is the overuse of
the target form on which learners received instruction or feedback (Henshaw, 2011; Sanz & Morgan-Short, 2005). A comparison between the WCF and No-WCF groups showed that the learners that received feedback produced considerably more cases of oversupply of the subjunctive. Moreover, the results also revealed that there were more participants in the WCF group (44.4%) that made errors of overuse as compared to the No-WCF (27.8%). These findings suggest that, not only did the treatment not help the WCF produce fewer errors in contexts were the subjunctive was obligatory, but it is possible that the information obtained from the feedback negatively affected their developing grammar in terms of the indicative-subjunctive distinction. Nonetheless, future research should examine to what extent the overuse that may result from form-focus interventions is persistent or whether it is restructured over time.

Comparison of Revision of the Present Subjunctive Errors

As expected, a comparison between the two groups in terms of their ability to successfully self-correct their errors showed that the participants who received WCF were able to revise significantly more errors than their counterparts in the No-WCF group, as in Example 7.

Example 7

(WCF)S17  *No pienso que el dinero da (de) la felicidad*  (Session 6)

‘I do not think that money *brings [indic.] (brings-subj) happiness’

In the example above, a participant in the WCF group that produced the present indicative form *da* ‘gives’ (-3rd person, -indicative) in a context where the subjunctive form *dé* ‘gives’ (-3rd person, -subjunctive) was required was able to successfully self-correct his/her error during the revision session after receiving WCF feedback. Although some learners in the No-WCF group were able to revise some of their errors without any external feedback, as in
Example 8 below, the results showed that the percentage of correct self-revisions was much higher in the feedback group (72%) than in the no-feedback condition (9%).

Example 8

(No-WCF)S9 No creo que el alcohol es (sea) sano (Session 8)

‘I do not believe that alcohol *is [indic.] (is-subj) healthy’

There are two main implications that stem from these findings. First, the fact that most of the errors that the learners produced were successfully revised indicates that the treatment, namely, the combination of indirect correction (i.e., boldface) and metalinguistic information (i.e., indication of the type of error that was made) was effective in assisting with the revision of present subjunctive errors. It should be also noted that the WCF was provided consistently since the majority of the subjunctive errors made by the learners were marked for them. On the other hand, the results suggest that the lack of feedback or indication that a present subjunctive error was produced in the output led to the learners’ inability to revise their errors, which is again consistent with what was found with the other linguistic forms that were examined in previous studies on error revision (e.g., Ashwell, 2000; Fathman & Whalley, 1990; Ferris, 1997; Ferris & Roberts, 2001).

Comparison of the Accurate Production of the Present Subjunctive

Similar to the analysis conducted with non-canonical gender marking, the production of individual forms in the present subjunctive were traced over time since, as argued earlier, grouping all of the forms into one single category (i.e., ‘present subjunctive’) may obscure instances in which learning may have or may have not occurred and would not provide clear evidence which with to answer some important questions, including the extent to which error
revision leads to the correct use of the same form or structure in new texts or whether the effect of the feedback maintained over time.

In order to provide a more detailed picture of the effects of the WCF treatment, the production of present subjunctive forms were classified according to whether the same revised form/structure was (a) immediately used by the same learner in the following text, (b) used by the same learner in a later but not immediate text, or (c) used in two (or more) different texts by the same learner.

The results, in terms of the accurate production of subjunctive forms in new texts, provided some interesting findings that will be discussed. First, with respect to the effect of error revision, the results showed that irrespective of the group condition when a present subjunctive error was not revised, the same type of error was produced in subsequent production, as in Example 9.

**Example 9**

(No-WCF)C2 *No creo que el dinero da la felicidad* (Session 6)

‘I do not think that money *brings [*indic.]* happiness’

(No-WCF)C2 *No creo que existe* (Session 7)

‘I do not think that there exists [*indic.*]’

The fact that the errors that were not revised were produced incorrectly in new pieces of writing could be anticipated for various reasons. First, it has been pointed out that if learners make an error, attention to that error is what allows them to become aware of a mismatch or gap between what they produce and the target form (e.g., Ellis, 1994; Gass, 1988; Schmidt & Frota, 1986; Swain, 1995). Hence, it is self-evident that if learners are not aware of the errors in their interlanguage, they will most likely produce the same non target-like forms in the future, as in
Example 9 above. For this reason, some authors (e.g., Long, 1996; Schmidt & Frota, 1986) have argued that an intervention that prompts attention to form, such as WCF, might be necessary to facilitate their interlanguage development.

Regarding the effect of error revision with and without WCF, it was found that, regardless of whether the revised forms were produced again in immediate texts, as in Example 10, or in later texts, as in Example 11, the target-like forms in the subjunctive were never incorporated correctly by the learners in the WCF and No-WCF groups.

Example 10

(No-WCF)S6  Es posible que hay (haya)  (Session 9)

‘It is possible that there *is [indic.] (is-subj)’

(No-WCF)S7  Es muy importante que toda la gente tiene  (Session 10)

‘It is very important that all of the people *have [indic.]’

Example 11

(WCF)S10  No pienso que los estereotipos son (sean) justos  (Session 2)

‘I do not think that stereotypes *are [indic.] (are-subj) fair’

(WCF)S10  No pienso que el dinero da la felicidad  (Session 6)

‘I do not think that money *brings [indic.] happiness’

In Example 10, a participant from the No-WCF group was able to detect that the indicative form hay ‘there is’ was incorrect and replaced it with the correct present subjunctive form haya ‘there is’; however, the same learner subsequently produced the indicative form in a context where the subjunctive form tenga ‘has’ was required. Example 11 illustrates the case of a learner who, after correctly replacing the non-target like indicative form son ‘are’ with the correct subjunctive form sean ‘are’, produced the same error type in a later text using the
indicative form *da* ‘brings/gives’, instead of the target form in the subjunctive. The combined results indicate that, in spite of being able to provide the correct form during the revision sessions, the learners always failed to produce the target form in a similar context. The fact that the error revision had no effect on the accurate production of the present subjunctive is in line with Collentine (2010) who found that instructional interventions are very limited in promoting the correct use of the subjunctive because of its multifaceted nature involving morphology, mood distinction, complex syntax and modality.

Apart from the complex nature of the subjunctive and mood acquisition, there are other reasons that may account for the inefficacy of the WCF treatment. If we transfer the basic notion of the Teachability Hypothesis (e.g., Pienemann, 1989; 1998) to the use of corrective feedback, that is, that “the influence of teaching is restricted to the learning of items for which the learner is ‘ready’” (Pienemann, 1989, p. 63), it could be argued that WCF, as it is typically implemented, can only promote acquisition when it targets linguistic forms that are learnable at a given point in time. In the case of the present subjunctive in Spanish, Collentine (e.g., 1995; 2010), while not directly citing the Teachability Hypothesis, argued that English L1 Spanish L2 learners cannot select mood in subordinate clauses until they can reliably produce subordinate clauses. Once learners reach that stage, they begin to incorporate the subjunctive into the verbal system, which is seen as another “conjugation” without any particular communicative value. As L2 learners continue acquiring the subjunctive, they start accumulating lexical and semantic features. With this in mind, the available data suggest that while the learners were at the complex syntactic stage, as shown by their consistent use of subordination, most of them were not yet at the mood selection level in production. The production of subordinate clauses primarily included the use of
the indicative and, to a lesser extent, the infinitive form in contexts were the target subjunctive was required as shown in Example 12 below.

Example 12

(WCF)S9  Quieren que todo el mundo a tener lo mismo  (Session 9)
They want that everyone *to have [inf.] the same
‘They want everyone to have the same’

(WCF)S19  No son justos porque es posible que causan  (Session 10)
They are not fair because it *is [indic.] possible that they cause
‘They are not fair because it is possible that they cause’

All things considered, there are several aspects of the treatment that could help explain why it was seemingly ineffective in facilitating the accurate production of the present subjunctive forms. The first factor has to do with the nature of the WCF and the extent to which it enabled the cognitive processes hypothesized to help with acquisition. Based on Schmidt’s (2001) distinction between noticing (e.g., attention to a not target-like form in the output) and metalinguistic awareness, it is possible that the treatment, which explicitly indicated the type of error, only increased noticing, in the form of error revision, but did not encourage a deeper level of cognitive processing necessary to acquire a complex form such as the present subjunctive, which requires morphological, syntactic, semantic and pragmatic knowledge, among others. Another factor that could have influenced the intake of the feedback is the learners’ knowledge of the subjunctive. In this regard, considering that the learners’ developmental stage in the acquisition of the present subjunctive, it is possible that the learners simply relied on their morphological knowledge to revise their errors without any further processing of the remaining linguistic information required to learn mood selection or the specific contexts in which it is
used. As a result, it could be argued that since the type of information that the learners obtained from the feedback is primarily morphological (i.e., that the indicative form needed to be replaced with the subjunctive), the remaining information (i.e., semantic, pragmatic, etc.) necessary to acquire mood selection was not likely to be processed or inferred due to their current level of development. Hence, all these factors, namely, the complexity of the acquisition of the subjunctive and mood selection among L1-English L2-Spanish learners, the current stage of acquisition, in addition to the limitation of the WCF to deliver linguistic information beyond the attention to error form, can describe why the participants in this study, regardless of self-correcting their errors, were not able to produce the same forms accurately in new texts.

5.3. Comparison between L2 and HL learners within and across groups

In order to answer the second research question, the error production, revision and accuracy development of canonical gender marking, non-canonical gender marking, omission of definite articles in obligatory contexts and present subjunctive were compared between the L2 learners and HL learners. Given that no prior study has investigated the effects of WCF feedback with HL learners, a null hypothesis was established for this question.

5.3.1 Canonical Gender Marking

Production of Canonical Gender Marking

Similar to the results of the WCF and No-WCF groups, the mean production and distribution of canonical gender marking production was comparable between the L2 and HL learners. As was mentioned earlier, the higher frequency of canonical gender assignment as compared to gender agreement was attributable to greater occurrence of the determiner-noun concord in Spanish, and also to the type of written genre that was employed in this study,
argumentation, which did not elicit the production of noun-adjective concord, as compared to other genres.

Error Production

First, a comparison between the HL and L2 learners in terms of error production and overall accuracy showed that, regardless of the group condition, the HL learners were more accurate and produced fewer gender marking errors than their L2 counterparts. These findings from data collected in an authentic classroom setting differed from those found in experimental research (Alarcon, 2011; Montrul et al., 2008), which showed that L2 learners performed better on written tasks compared to HL learners. These studies argued that the reason why L2 learners generally performed better in writing was because they started to acquire the language in formal classroom settings in which there is heavy emphasis on reading, writing, and the development of metalinguistic skills, in contrast to HL learners, who initially learned the language verbally, and later on received instruction in the heritage language. However, while these differences might be the case in lower level courses or in L2 and HL learners of lower proficiencies, the reality is that in more advanced language courses where both HL and L2 learners are mixed, as in the one from this study, the HL learners are typically more proficient both orally and in writing.

Furthermore, when comparing the participants’ experience with language instruction, the HL learners in this study reported that, similar to their L2 counterparts, they were enrolled in Spanish courses in either middle school or high school. These findings are not surprising considering that most schools in the United States offer Spanish as a foreign language, and they also serve to attest that any potential differences in written accuracy between these two types of learners due to L2 literacy or instruction in the L2 should be ruled out. Hence, differences in performance between the L2 and HL learners in this study might be better explained in terms of
age of acquisition and exposure to Spanish. In contrast to the L2 learners, the HL learners were exposed to Spanish since birth and even if they underwent language attrition or incomplete acquisition due to a decrease in input later on (Montrul, 2008), their linguistic experience may have rendered them a significant advantage over their L2 counterparts in terms of the acquisition of the canonical gender feature, which resulted in the production of fewer canonical gender marking errors.

With respect to the L2 learners in the WCF and No-WCF groups, the results found no significant differences in the total number of gender marking errors between the two conditions. Likewise, the total number of errors produced by the HL learners in the WCF and No-WCF groups was comparable. Considering that the participants in this study were enrolled in the same course and presented similar canonical gender marking accuracy results at the beginning of the data collection period, the combined results provide preliminary information regarding the limited effect of the treatment in reducing the overall number of canonical gender marking errors at the group level.

Error Revision

In order to test the effects of the WCF and No-WCF conditions on error revision, the number and proportion of canonical gender marking errors revised by the L2 and HL learners were compared within and across the two groups.

With respect to the feedback group, the results showed that L2 learners received more corrections on both canonical gender assignment and agreement errors than the HL learners. These differences in the number of corrections were likely due to a higher occurrence of canonical gender marking errors in the production of the L2 learners as compared to that of the HL learners. In other words, the HL learners in the WCF group were more accurate than their L2
counterparts, and as a result, received fewer corrections. In terms of the learners’ ability to revise their gender marking errors, the results of this study are consistent with previous studies on error revision (e.g., Ashwell, 2000; Ferris & Roberts, 2001; Sachs & Polio, 2007), as it was found that those learners who received indirect WCF for the errors were able to make more accurate revisions than those who did not receive any CF. Nonetheless, there are some differences between the L2 and HL learners that are worth considering. First of all, when comparing the L2 and HL learners in the feedback group, the percentage of successfully revised canonical gender assignment errors was considerably higher in the L2 learners (84.4%) than in their HL counterparts (57.1%), while, on the other hand, no differences were found in the percentage of gender agreement errors successfully revised by the L2 learners (82%) and the HL learners (75%). Hence, the question that arises is why the feedback treatments were not as effective in helping HL learners successfully revise their canonical gender assignment errors. One possible reason may relate to some of the potential problems of comprehensive feedback identified by Sheen et al. (2009), including the limited time that learners had to revise all of their errors and the fact that some learners may decide to focus on certain errors at the expense of others.

Another factor that needs to be regarded is the frequency of canonical gender marking errors and the number of corrections that the HL learners received. The results showed that the L2 learners in the WCF group produced significantly more assignment errors and, consequently, received more corrections than their HL counterparts. Differences in the amount of feedback received might have contributed to make canonical gender marking errors more noticeable in L2 learners’ output and, as a result, they were more prompted to revise those errors, as compared to their HL counterparts.
Regarding the L2 and HL learners in the No-WCF group, the results showed that the HL learners were more accurate overall in their production of canonical gender marking than their L2 counterparts; however, in terms of error revision, they were only better at revising canonical gender agreement errors without the help of feedback. Hence, given that the HL learners were considerably more proficient in gender marking, one could have expected that this knowledge would have also made a difference for them to notice and correct their gender marking errors when compared to the L2 learners in the no-feedback condition. Again, a possible explanation refers to the low frequency of canonical gender assignment errors compared to gender agreement errors in the production of the HL learners, which could have made them less noticeable or salient. Moreover, it is also possible that the HL learners might have focused on certain specific errors during the revision sessions at the expense of others. In this sense, a closer look at the most frequent error types revised by the HL learners in the No-WCF revealed that their primarily focus was accent marking, by means of adding a diacritical mark to a word, with an average of 11 revision of this type per participant (SD=14) across all sessions, in addition to changes related to content (i.e., expanding or modifying an idea or information about the topic), with an average of 8.33 revisions per participant (SD=9.8). These results accounted for 38% and 29% of all of their revisions, respectively, which provide an indication that the HL learners actually centered their attention on some particular error types instead of others, which, in turn, may help explain why some errors, such as canonical gender marking, were overlooked or ignored.

Furthermore, a comparison between the HL and L2 learners in the No-WCF group, revealed that the L2 learners provided fewer content changes (M=5.92; SD=5.5), which accounted for 16% of all of their revisions, while their most frequently revised errors involved spelling (M=8.17; SD=7.1) and word choice (M=7.5; SD=5.1), which comprised 22% and 20%
of all of their revisions, respectively. These findings add to Torres’s (2013) who found in a study with recasts that, compared to L2 learners, HL learners tend to focus more on content as compared to form, which may be related to their experience with the heritage language, in communicative contexts, whereas the L2 are typically exposed to metalinguistic information.

In conclusion, as hypothesized, the comparison across learners and conditions revealed that the L2 learners that received WCF were able to revise significantly more gender marking errors that the L2 and HL in the no-feedback condition. These findings are in line with previous studies (e.g., Ferris & Helt, 2000; Lalande, 1982) that have claimed that indirect corrections, by which students are informed that an error existed in their output but the correct form is not provided, are effective in helping students revise their errors. Additionally, the comparison across groups also showed that in the absence of WCF, most L2 and HL learners were not able to detect and revise their canonical gender marking errors, which makes the case for the use of indirect CF as an instrument to facilitate error noticing and revision of this type of error.

**Canonical Gender Marking Accuracy Development**

A comparison between the HL and L2 learners in terms of accurate production of canonical gender marking has shown differences both across and within the two group conditions that need to be considered.

First of all, in terms of the accurate use of canonical gender marking, similar to what Montrul et al. (2008) found with low-intermediate proficiency L2 and HL learners in oral production, the written production results in this study showed that gender concord was more accurate for determiners (i.e., gender assignment) than for adjectives (i.e., gender agreement). Differences in accurate production between these two types of canonical gender marking forms are not surprising due to the inclusion of both attributive and predicative adjectives in the gender
agreement count, which, according to the literature on gender marking acquisition (e.g., Bruhn de Garavito & White, 2003), are typically acquired later than gender assignment.

In terms of canonical gender marking accuracy development, the results revealed differences between L2 and HL learners both within and across conditions in both canonical gender assignment and agreement. With respect to the WCF group, the results showed that the L2 learners experienced a slight increase in the accurate production of canonical gender assignment while the accuracy of the HL learners slightly decreased over time. If, as hypothesized, we assume that error noticing leads to acquisition, the fact that the quantity and frequency of assignment errors revised by the L2 learners was much higher than their HL counterparts might explain to some extent the efficacy of the treatment and why the L2 learners exhibited an increase in accuracy over time. The relationship between error revision and accuracy development appear to be further supported by the L2 learners in the No-WCF group who provided very few revisions of their errors and experienced a decrease in the accurate production of canonical gender assignment over time. With respect to the HL learners in the No-WCF group, their linear trend revealed the most accurate performance of all of the groups in the production of canonical gender assignment with scores at ceiling or near-ceiling. These results suggest that the majority of the HL learners in the No-WCF group might have already acquired the canonical gender assignment feature, and, consequently, the use of CF for the purpose of acquisition would have been unnecessary. When comparing the HL learners across the two conditions, while the production of canonical gender assignment in the HL learners was highly accurate with over 95% of correct use, their developmental patterns revealed a small decline over time. Whether these results were due to the limited amount of feedback received by the HL learners in the WCF group or their decision to focus on certain error types and ignore the
feedback, it is clear that the treatment did not have an effect on their accuracy development of canonical gender assignment.

With respect to the accuracy of canonical gender agreement, as hypothesized, the HL learners in this study were more accurate than their L2 counterparts. The observed trends showed that the HL learners in the No-WCF group were the most accurate in the production of canonical gender agreement, followed by the HL learners in the WCF group; while the L2 learners in the WCF group and No-WCF, similar to what was found for gender assignment, scored below their HL counterparts in canonical agreement.

Regarding the effects of the treatment on the accurate production of canonical gender agreement, the L2 learners in the WCF group exhibited a moderate decline in accuracy, while the accuracy level of their HL counterparts remained stable over time. These findings provide a clear indication that the WCF treatment did not have a positive effect on the accuracy development of this linguistic feature, especially when compared to the L2 and HL learners in the control group. The patterns showed no variation in the accurate production of the L2 learners in the No-WCF group, while their HL counterparts were able to improve their production and reach 100% correct use of canonical gender agreement. Hence, one of the main questions is why the L2 learners in the WCF group, who revised significantly more gender agreement errors than any of the other participants in the two conditions, did not improve their accuracy but, instead, experienced a decrease in accuracy over time. In this sense, there are several explanations that may account for this situation. Some authors (e.g., Sheen et al., 2009) have argued that when learners are exposed to the correction of a variety of grammatical features, like the participants in this study, they might have more problems to process different error types at the same time and also retain the feedback effectively. However, a further question that results from this is why the
L2 learners in the WCF improved their accuracy in canonical gender assignment but not in gender agreement. One reason could be the fact that the L2 learners received and revised a larger number of gender assignment errors (N=38) as compared to gender agreement errors (N=23), which may have led to greater attention to that particular feature, for noticing and subsequent learning. Another explanation relates to one of Truscott’s main arguments against the use of WCF on the basis that the acquisition of grammatical structures is a complex and gradual process and “not a sudden discovery as the intuitive view of correction would imply” (Truscott, 1996, p. 342). According to Truscott, WCF involves a simple transfer of information that may enable learners to correct their errors, but as it is typically implemented, fails to acknowledge the complex nature of interlanguage development. One of his main arguments is that providing WCF at a time when the learner is not “ready” to acquire a particular linguistic feature, that is, a focus on forms or structures that are out of sequence of the natural order (e.g., Krashen, 1985), or outside their ‘zone of proximal development’ (ZPD) (e.g., Aljaafreh & Lantolf, 1994) will not foster true accuracy development, but instead “pseudo-learning” or revision skills at best. In contrast to this claim, Bitchener and Ferris (2012) argued that the learners who are ready to acquire the form or structure targeted by written CF may show greater accuracy over time in the use of the targeted linguistic form or structure. However, while the notions of “learner’s readiness” or ZDP are frequently used in SLA theory to help explain why learners might (or might not) acquire certain L2 forms or be responsive to certain input, the fact is that it is very difficult to operationalize those notions. Hence, it is reasonable that looking at the accuracy development of partially acquired forms that are already part of the learners’ interlanguage, such canonical gender marking, will provide solid evidence on the efficacy of WCF. In this sense, the
results showed that the provision and revision of multiple canonical gender agreement errors, in addition to plenty of opportunities for production, made no difference in terms of acquisition.

Additionally, it should be noted that, while timed written production, as opposed to oral tasks, is not regarded as the most representative measure of implicit knowledge (Ellis, 2005), still, the written data collected for this study can offer valuable information on the learners’ ability to access their linguistic knowledge under time constraints and, thus, provide insight into the question of whether the gender feature can be acquired by L2 learners of Spanish. In this sense, the gender assignment results offer some interesting findings that indicate that L2 learners, especially those in the WCF group, were able to improve their accurate production over time, which suggest that they may eventually achieve native-like performance if they continue with the same progression. Moreover, although there is no previous research that have identified patterns of gender marking in the free-written production in L2 learners of Spanish, the accuracy results in this study are in line with those from Alarcon (2011) whose L2 learners by their fourth semester of college Spanish obtained 95% correct gender assignment on her written tests intended to measure gender marking acquisition. Similarly, White et al. (2004) found that intermediate and advanced groups of L1 English L2 Spanish performed above 90% accuracy on gender marking in oral production and written comprehension tasks, which can be compared to the results obtained by the L2 learners in this current study in free-writing production.

Furthermore, a comparison between the two types of canonical gender marking revealed that both the L2 and HL learners were less accurate in the production of canonical gender agreement as compared to gender assignment, regardless of the group condition. These findings are in line with previous research (e.g., Fernandez, 1999; Franceschina, 2005; Hawkins & Franceschina, 2004; McCarthy, 2007) that found that L2 learners of Spanish performed better
with gender marking with articles than adjectives, leading them to conclude that gender assignment with articles is acquired earlier than gender agreement with adjectives. Likewise, the patterns found in the writing of HL learners in this study showed higher accurate production of canonical gender assignment than gender agreement. These findings obtained through free-written production are also consistent with previous experimental studies on canonical gender marking (e.g., Montrul et al., 2013; Montrul, et al., 2008) which found higher performance in gender assignment among HL learners of different ages and proficiency levels.

5.3.2. Non-Canonical Gender Marking

Production of Non-Canonical Gender Marking

The results showed that the average production and distribution of non-canonical gender marking was comparable between the L2 and HL learners within and across the two conditions, including a significantly higher occurrence of non-canonical gender assignment as compared to gender agreement. These findings are similar to the ones observed with canonical gender marking and suggest that mode of language acquisition (L2 vs. HL) did not have an effect on the production of non-canonical gender marking, but rather, as argued earlier, the differences are likely due to the high occurrence of the Spanish NP requiring determiner-noun concordance (i.e., gender assignment) in addition to the type of written genre, argumentation, which did not prompt the production of non-adjective concordance (i.e., gender agreement).

Comparison of Non-Canonical Gender Marking Error Production

The results of this study provide some interesting information regarding instances of non-canonical gender marking errors in the written production of the L2 and HL learners within and
across the two conditions, and the extent to which their errors were noticed and self-corrected during the revision sessions.

First, both the L2 and HL learners produced comparatively more errors with non-canonical ending nouns than with canonical nouns. These results were expected considering that research has shown that both L1 and L2 learners require more time to learn and process non-canonical nouns compared to canonical ending nouns. For example, Hernandez-Pina (1984) in her case study on L1 acquisition, and Montrul (2011) in a study of L1 attrition with a Guatemalan adoptee, reported that most gender errors occurred with non-canonical nouns. Moreover, studies with L2 learners and HL speakers of Spanish (e.g., Alarcon, 2011; Montrul et al., 2008, 2013, 2014) have found that both L2 learners and heritage speakers were more inaccurate with non-canonical nouns than with canonical gender ending nouns in various tasks conducted in experimental settings, which included written comprehension, oral production and grammaticality judgment tasks. Several explanations may account for the variation found in the accurate production of canonical and non-canonical gender marking, including differences in terms of acquisition, frequency in the input, and use. Canonical ending nouns follow a regular morphological pattern in Spanish through the inflectional markers -o masculine and –a feminine. According to the dual mechanism model of inflection (e.g., Pinker 1999; Pinker & Ullman, 2002), regular morphological processes, such as canonical gender marking in Spanish, are stored in procedural memory and, when learned, become automatized. In contrast, the gender of non-canonical ending nouns, which are also less frequent in the input, need to be memorized due to their irregular patterns (-e, consonant, opposite vowel). For these reasons, Montrul et al. (2014) suggest that reduced input and use of Spanish by L2 learners and heritage speakers may affect
the storage in declarative memory and the strength of the lexical association of non-canonical ending nouns resulting in higher frequency of errors.

With respect to the distinction between gender assignment (1) and agreement (2), errors of these two types of non-canonical marking were found in the written production of the L2 and HL learners, regardless of the condition.

(1) L2  
*Y una programa publica*

‘And a [*fem]* public [*fem]* program [*masc]*’

HL  
*Es muy importante en muchos partes*

‘It is very important in many [*masc]* parts [*fem]*’

(2) L2  
*Las acciones sencillos*

The [*fem]* easy [*masc]* actions [*fem]*

HL  
*De la union europeo es diferente*

‘Of the [*fem]* european [*masc]* union [*fem]* is different’

Based on previous work (e.g., Alarcon, 2011; Montrul, 2008, 2013), it was expected that both the L2 and HL learners in this study would be less accurate with gender agreement with non-canonical ending nouns, as compared to gender assignment, given that gender assignment with articles is acquired earlier than gender agreement with adjectives (e.g., Fernandez, 1999; Franceschina, 2005; Hawkins & Franceschina, 2004; McCarthy, 2007). Furthermore, a comparison between the L2 and HL learners in terms of non-canonical gender marking revealed that overall the HL learners produced fewer errors during production than their L2 counterparts in both gender assignment and agreement, regardless of the group condition. These findings are again in line with the results reported for canonical gender marking. As mentioned earlier, some of the main variables that may have given the HL learners an advantage over L2 learners in
terms of performance are those related to age and context of acquisition and linguistic experience, particularly with oral production. In this sense, the L2 learners in this study indicated no use of Spanish outside the classroom, while their HL counterparts reported over fifty percent of use of Spanish at home (M=51.7; SD= 30.1). In terms of non-canonical marking knowledge, these differences in language use and experience, as Montrul et al. (2014) suggested, may have affected their retrieval from declarative memory that resulted in errors of production. Moreover, when comparing the total number of non-canonical gender errors across groups and participants, the results showed no differences between the L2 learners in the WCF and No-WCF groups and between the HL learners in the two conditions. However, it should be noted again that a measure grouping all non-canonical gender marking errors does not provide evidence of the effects of the treatment, because in many cases the non-canonical ending nouns and errors made by one participant were different from the ones produced by others, and, consequently, cannot be compared.

Non-Canonical Gender Marking Error Revision

In terms of the learners’ ability to revise their errors with and without WCF, a comparison between the L2 and HL learners within and across conditions revealed some similarities and differences that are worth discussing.

First, regarding the WCF group, the results showed no differences in the total number of corrections of non-canonical gender marking that the L2 and HL learners received. However, a closer look showed that, although all the L2 learners in the WCF group received some error corrections for non-canonical gender marking errors, only half of the HL learners made gender errors and, consequently, received WCF. These results indicate that, while the number of corrections was comparable between the L2 and HL learners in the WCF, the corrections were
more spread among the L2 learners and concentrated in just a few HL learners as the rest were completely accurate in the production of non-canonical ending nouns.

With respect to the learners’ ability to revise their non-canonical gender marking errors, there were no differences between the L2 and HL learners in the WCF in terms of the proportion of errors that they were able to successfully revise. Likewise, the results in the No-WCF group showed that the L2 and HL learners did not differ in the number and distribution of non-canonical gender marking error revision. However, it is important to note that, in the case of the No-WCF group, the L2 learners produced more non-canonical gender marking errors than their HL counterparts. Furthermore, a comparison across conditions indicated that the total number and proportion of errors successfully revised by the L2 and HL learners in the WCF group was significantly higher than their counterparts in the No-WCF group. The combined results further confirm that error revision with indirect WCF was more effective than revision without feedback to help both L2 and HL learners notice and self-correct their non-canonical errors. However, it should be noted that, comparatively, the proportion of non-canonical gender marking errors that the learners were able to revise with the help of feedback was much lower than that observed with canonical gender marking. A possible reason for this could be the fact that, unlike direct feedback, which involves providing correct revisions/forms for the students, indirect WCF only offers hints (e.g., underline, circles, etc.) and relies on the students’ ability to come up with their own correct answer. As such, indirect WCF might have been more effective with canonical marking because of its overt morphological expression (ending – o for masculine and – a for feminine nouns) and this may have served as a visual cue when the learners were trying to identify their concordance errors. Indirect WCF, on the other hand, might have been less efficient with non-canonical errors, given that learners could not resort to a consistent rule in
addition to the lack of transparency of the irregular markers for masculine and feminine (ending -e, consonant, opposite vowel), making them more difficult for learners to notice in the output. The same arguments could also explain why the L2 and HL learners in the no-feedback condition were able to provide more revisions of canonical than non-canonical gender marking errors.

Canonical Gender Marking Accurate Production

As mentioned earlier in this chapter, the effects of the WCF treatment on the accurate production of non-canonical gender marking in L2 and HL learners was measured by tracing the forms that were previously revised (or not revised) and produced in new pieces of writing. While the production of the L2 learners contained some errors that could be traced over time, the HL learners, especially those in the No-WCF group, were very accurate with non-canonical noun ending concordances, and, as a result, their output did not include sufficient traceable forms to allow for a comparison with the L2 learners. While the weaker links hypothesis (Gollan, Montoya, Cera & Sandoval, 2008) suggests that noun-gender links of the HL speakers may have weakened progressively as their first language became the secondary language, the results indicate that the lexical association links of HL learners in this study remained very strong, at least for those nouns that they chose to use in their written production.

The idea of language use and frequency is further supported by the fact that the HL learners, as indicated earlier, reported the use of Spanish outside of class on a daily basis, which explains why their lexical associations for non-canonical ending nouns might not have been affected under L1 attrition. In the same vein, the reduced frequency of language use outside the classroom would help explain why the L2 learners, by contrast, were never able to produce certain non-canonical gender marking correctly, including forms for which they received indirect
WCF. In fact, the findings on the accurate production of non-canonical assignment forms that were previously revised further support the view that the knowledge obtained from the error revision and WCF was not maintained over time. The results showed that the L2 learners were very accurate when using a revised non-canonical ending noun in immediate production (80%), as with oportunidades ‘opportunities’ (fem.) in Example 1 below.

Example 1

(WCF)S15  *Tienen el mismos (las mismas) oportunidades*  (Session 4)
‘They have the [masc]* same [masc]* (the [fem] same [fem]) opportunities’

(WCF)S15  *La tecnologia abre la puerta a muchas oportunidades*  (Session 5) 
en el mundo. Las oportunidades incluyen la comunicación
‘Technology opens the door to many [fem] opportunities [fem] in the world. The [fem] opportunities [fem] include communication’

However, the correct use of non-canonical ending nouns decreased drastically when the same revised form was used in subsequent texts (10.7%), as in Example 2 países ‘countries’ (masc.) and also when the same non-canonical form was used for the second time in a new piece of writing, as in problemas ‘problems’ (masc.) in Example 3.

Example 2

(WCF)S16  *Cambia a la cultura de otras (otros) paises*  (Session 9)
‘It changes the culture of other [fem]* (other [masc]) countries [masc]’

(WCF)S16  *Es cuando personas de otras paises*  (Session 11)
‘It is when people from other [fem]* countries [masc]’
Example 3

(WCF)S9  
\textit{El dinero no para las (los) problemas}  
(Session 6)

‘Money does not stop the \textit{[fem]}* (the) \textit{[masc]} problems \textit{[masc]}’

(WCF)S9  
\textit{Yo creo que hay un problema}  
(Session 7)

‘I think there is a \textit{[masc]} problem \textit{[masc]}’

(WCF)S9  
\textit{Pero la problema es que}  
(Session 11)

‘But the \textit{[fem]}* problem \textit{[masc]} is that’

As Montrul et al. (2014) noted, reduced frequency of use leads to slower retrieval of nouns in the lexicon and slower speeds at computing syntactic dependencies, including concord with determiners, nouns and adjectives, which all lead to gender marking errors. All things considered, the implication of this notion in addition to the findings of this section is that the long-term effects of feedback with low-frequency forms and structures, such as non-canonical ending nouns, is likely to be fairly limited in the foreign language classroom because most of the time L2 learners do not have the opportunity to practice and retrieve those forms regularly in order to activate their declarative knowledge and reinforce their lexical associations.

5.3.3 Definite Article Omission

This research question was aimed to compare the L2 and HL learners with respect to the effects of the WCF on the production, error revision and accurate written production of definite articles in Spanish.

\textit{Production of Definite Articles in Obligatory Contexts}
The results showed that the occurrence of definite articles in obligatory contexts was comparable between the L2 and HL learners within and across the two conditions, which ranged from an average of 4.4 cases per text for the L2 learners in the WCF group to 4.7 cases for the HL learners in the WCF group. The fact that all of the participants answered the same prompts may explain why there were no differences in terms of the rate of generic nouns that required the use of definite articles. Some of these nouns that elicited the use definite article and were frequently found in the writing of both the L2 and HL learners included, *hombres* ‘men’, *mujeres* ‘women’, *personas* ‘people’, *dinero* ‘money’, *globalización* ‘globalization’ or *religión* ‘religion’.

**Error Production**

A comparison in terms of the total number of errors of definite article omission for the L2 and HL within and across the two conditions revealed some differences that are worth discussing. First of all, the production results showed that both L2 and HL learners made errors of omission, irrespective of the condition, as in Example 1 with the generic nouns *globalización* ‘globalization’ and *estereotipo* ‘stereotype’.

**Example 1**

(WCF) L2  *Globalización es el proceso de*  
⊕ [art omit]* Globalization is the process of

‘Globalization is the process of’

(WCF) HL  *Globalización casi siempre se refiere a países*  
⊕ [art omit]* Globalization almost always refers to countries

‘Globalization almost always refers to countries’

(No-WCF) L2  *Yo no pienso que estereotipos son justos*  
I do not think that ⊗ [art omit]* stereotypes are fair
‘I do not think that stereotypes are fair’

(No-WCF) HL  Estereotipos son mas populares con grupos raciales  

(Session 2)

Stereotypes are more popular with racial groups’

‘Stereotypes are more popular with racial groups’

These findings are in line with previous research that has reported a transfer effect with errors of article omission from English into Spanish in adult L2 acquisition and heritage speakers (Lipski, 1993; Montrul & Ionin, 2012).

Furthermore, with respect to the effects of the treatment on the overall error production, the results of this study indicated that the L2 learners in the No-WCF group (M=1.07) produced almost double as many errors of article omission as their L2 counterparts in the WCF group (M=0.53). Likewise, the HL learners in the feedback condition averaged fewer errors of article omission than their HL counterparts in the No-WCF groups across all sessions; however, the differences were not statistically significant. While these results, especially the L2 learners in the feedback condition, seem to be congruent with what WCF advocates have claimed, namely that frequent feedback, in addition to written practice, can lead to error reduction, there are other outcomes of the treatment that need to be considered. First, similar to gender marking, there was great variation in the total number of omission errors across all session for the L2 learners, ranging from 0 to 22 errors in the WCF, and from 3 to 31 total errors per participant in the No-WCF group. Likewise, the HL learners also showed considerable variation in the total production of errors of omission, which ranged from 0 to 21 in the WCF group and from 0 to 16 in the No-WCF group. These findings in addition to the small-medium size effect (d=0.18) found for L2 learners in WCF group suggest that not all of the learners responded equally to the treatment, which will be discussed later in more detail when comparing the effects of the
treatment on accuracy development. Another important question to consider is that some L2 and HL learners did not make any article omission errors, including 33.3% of the HL learners and 16.67% of the L2 learners. There are two important implications related to these findings. First of all, while it is not surprising that there were HL learners who did not make any article omission errors given their early language acquisition and linguistic experience, the fact that some L2 learners did not produce any omission errors suggest that they might have learned the generic interpretation with definite articles in Spanish, as in Example 2.

Example 2

(L2)S18  En el pasado, los mujeres no ganaban tanto como los hombres (Session 3)

In the past, the women did not earn as much as the men

‘In the past, women did not earn as much as men’

(L2)S18  El dinero crea un mundo aislado (Session 5)

The money creates an isolated world

‘Money creates an isolated world’

(L2)S18  Creo que el problem con el alcohol (Session 7)

I think the problem with the alcohol

‘I think the problem with alcohol’

(L2)S18  Por ejemplo, el cristianismo tiene (Session 10)

For example, the Christianity has

‘For example, Christianity has’

The excerpt in Example 2 shows the accurate production of an L2 learner with definite articles with generic interpretation, including general concepts el cristianismo ‘Christianity’,
nouns that refer to all members of a class los hombres ‘the men’ and las mujeres ‘the women’, and substances el alcohol ‘alcohol’. It should be also noted that the L2 learner in Example 2 correctly applied the generic rule to the noun mujeres ‘women’ [fem.], however the noun was assigned the incorrect gender los ‘the’ [masc.]. This is interesting because it could point a case of cognitive overload, in which the application of the generic rule did not allow the learner retrieve the correct gender.

Whereas these written samples may suggest that the learner was able to access and produce the generic use of the definite article in multiple contexts, given that all cases of definite articles in obligatory contexts were added to the count, it is not possible to really know the extent to which the information in the input helped them notice the generic interpretations with bare plurals or whether the prompts in the target language, as mentioned earlier, served as positive evidence. Another important theoretical issue has to do with the fact that HL learners produced fewer errors overall than their L2 counterparts, which raises the question of whether the differences in performance between these two groups were due to earlier acquisition or language proficiency. In this regard, while the age of acquisition factor suggests that effects of transfer may be stronger in L2 learners than in heritage speakers, a Pearson bivariate correlation between the learners’ proficiency scores before the treatment and their total number of errors across sessions revealed a significant negative correction between those two factors ($r = -0.59$, $p = 0.0002$). These results indicate that the higher the proficiency level, the fewer article omission errors, which is along the lines of Montrul and Ionin (2012) who claimed that language dominance matters more than age of acquisition for transfer.
The effects of the WCF treatment was measured by comparing the total number and proportion of errors of definite article omission correctly revised by the L2 and HL learners within and across the two conditions.

Regarding the WCF group, the results showed that most of the article omission errors made by the L2 and HL learners were successfully self-corrected during the revision sessions, as in Example 3.

Example 3  
(L2)S4 Los estereotipos causan que (Session 2)  
(article omission) (las) personas piensan en  
The stereotypes cause that ☐ [art omit]* (the) people think about  
‘Stereotypes cause people to think about’
(HL)S3 Cambia la manera en que piensan (Session 2)  
(article omission) (las) personas  
It changes the way in which ☐ [art omit]* (the) people think  
‘It changes the way in which people think’

In Example 3, the L2 and HL learners received WCF for the omission of the definite article las ‘the’ (fem. pl.) before the generic noun personas ‘people’ and successfully revised them by adding the correct target form. These findings indicate that the indirect WCF was very effective in helping both L2 and HL learners revise their definite article omission errors and they also point to the consistency of the treatment as most of the errors that the learners made were marked for revision. Again, one of the factors that could have contributed to the efficacy of the treatment is the information included in the feedback, which included the explicit indication of the error type and the location. Conversely, the results also showed that the L2 and HL learners
in the No-WCF group did not revise most of their definite article omission errors in their texts during the revision sessions, as in Example 4.

**Example 4**

(L2)S5  \textit{Religión ayuda personas con} \textit{[art omit]}  
\textit{Religion helps people with}  
‘Religion helps people with’

(HL)S7  \textit{Religion también es} \textit{[art omit]}  
\textit{Religion is also}  
‘Religion is also’

In Example 4, the L2 and HL learners omitted the definite article \textit{la} ‘the’ (fem. sg.) before the generic noun \textit{religión} ‘religion’ and did not revise these errors during the subsequent revision session. These findings suggest that, when feedback is not provided, the learners are not able to detect and revise their definite article omission errors.

**Accuracy Development of Definite Articles in Obligatory Contexts**

The accuracy trends for the use of definite articles in generic contexts indicated that the accurate use of definite articles in generic contexts increased over time in the L2 and HL groups in the WCF condition. On the other hand, the accuracy trend of the HL learners in the WCF slightly increased over time, while their L2 counterparts exhibited a small decline during the data collection period. The first implication of these findings is that for many learners the lack of indication that an error in the output existed led to recurrent erroneous production of the same form, as in Example 5.

**Example 5**

(L2)S4  \textit{Personas que no tienen un trabajo}  
‘Persons that do not have a job’
Proponents of the use of WCF have claimed that learners who do not receive any feedback, as a reactive way of form-focused intervention, like the students in Example 5, may develop fossilized non-target like grammar. In the sample above, the two learners produced the same error of article omission with the generic noun personas ‘people’ in different texts and never produced the correct form during the data collection period. However, it should be noted that definitions of fossilization vary widely and it is also very difficult to distinguish permanent fossilization from temporary stabilization of the IL, which can last for several years (Long, 2003).
With respect to the feedback condition, the results suggest that for some learners the WCF treatment led to improved use of definite articles before some generic nouns, as in Example 6.

**Example 6**

(L2)S7 *(article omitted) (las) personas necesitan ayuda* (Session 1)

∅ [art omit]* (the) People need help

‘People need help’

(L2)S7 *En realidad, las personas estan* (Session 6)

Actually, the people are

‘Actually, people are’

(L2)S7 *Las personas que beben alcohol* (Session 8)

The people who drink alcohol

‘People who drink alcohol’

In Example 6, an L2 learner, after receiving WCF and successfully self-correcting an article omission error, produced the correct form *las* ‘the’ before the generic noun *personas* ‘people’ in two different texts. However, the data also showed that the generic knowledge that the learners may have obtained from the feedback was not transfer to other types, as in the Example 7 below, where the same L2 learner omitted the article with the generic noun *globalización* ‘globalization’

(L2)S7 *Globalizacion es cuando un pais* (Session 11)

∅ [art omit]* Globalization is when a country

‘Globlization is when a country’
While the overall results may suggest that the L2 and HL learners benefited from the WCF as compared to the no-feedback condition, the individual patterns showed that most students (55.6%) in the feedback condition either exhibited a decrease in accuracy over time or their accuracy use of definite articles in obligatory contexts remained stable, as in Example 7.

**Example 7**

(L2)S9 *(article omitted)* (las) *Personas con problemas* (Session 1)

Ø *(art omit)* (the) People with problems

‘People with problems’

(L2)S9 *No todas* *(article omitted)* (las) *personas en un grupo* (Session 2)

Not all Ø *(art omit)* (the) people in a group

‘Not all people in a group’

(L2)S9 *Cuando* *(article omitted)* (las) *personas no tienen* (Session 5)

When Ø *(art omit)* (the) people do not have

‘When people do not have’

(L2)S9 *No todas personas necesitan* (Session 10)

Not all Ø *(art omit)* people need

‘Not all people need’

Example 7 illustrates the case of an L2 learner who never produced the correct definite article *las* ‘the’ before the generic noun *personas* ‘people’, in spite of receiving and successfully revising the same error multiple times. There are several factors related to the acquisition of the generic interpretation in L1-English L2 Spanish learners that may explain why learners who received WCF and also those who did not receive any feedback were not able to improve their use of definite articles before generic nouns. One of these factors has to do with the subset-
superset relationship with English as the superset and Spanish as the subset. Based on Slabakova’s (2006) argument, similar to Italian, learning the generic interpretation in Spanish is significantly challenging for L1-English L2 Spanish learners because they need to ‘unlearn’ one native interpretation, the generic one with bare nouns, with very little access to positive (or negative) evidence. As such, the poverty of stimulus (POS) situation and the Subset Principle (White, 1989) could contribute to the enduring non-native like state constrained by L1 transfer in the interpretation of generic nouns in L2 Spanish that result in errors of production.

Another factor to consider is related to the complex process of acquiring certain L2 forms and rules. According to Gelman and Raman (2003), in order to master the generic/specific distinction, children need to pay attention to at least three cues, including morphological cues, pragmatic cues, and world knowledge cues. In this sense, it remains unclear whether the information that learners obtained from revising their errors, as in Example 6, actually made them aware of the different interpretations or if it merely resulted in ‘pseudolearning’, that is, a superficial and temporary form of knowledge with little value for actual use of the language (e.g., Truscott, 1996, 1998). Hence, a possible way to test the learning effects of WCF, which should be considered in future research, is by complementing the analysis of longitudinal written production with experimental tasks designed to examine the extent to which the learners understand the different uses and interpretations of target forms.

5.3.4 Present Subjunctive

This question compared the extent to which comprehensive WCF helped both L2 and HL learners with respect to their error production, revision and whether their error revision led to subsequent accurate production of specific present subjunctive forms in new pieces of writing.

*Production of the Present Subjunctive*
As indicated earlier in this chapter, the production of the present subjunctive requires the knowledge of general morphological rules based on inflections in addition to the modality that is intended. For the purpose of this study, forms that were ambiguous or may accept both the indicative and subjunctive were not included in the count, as in Example 1.

Example 1

(WCF)HL  Aunque todos estamos viviendo en este país juntos,  (Session 3)
          seguimos teniendo nuestras tradiciones

‘Although we are [indic.] all living in this country together, we keep having our traditions’

In the above example, the adverbial clause is introduced by the conjunction aunque ‘although’, which may accept the indicative form estamos ‘are’ or the subjunctive ‘are’ estemos, depending on whether the speaker/writer is stating a fact or whether the outcome is unknown. Hence, only those instances in which the present subjunctive was obligatory were extracted and sorted according to the type of structure that elicited its use, including conjunctions of sentences (i.e., adverbial clauses), as in Example 2, the nouns that were modified (i.e., adjectival clauses), as in Example 3, and the verbs (i.e., nominal clause) in the main clause, Example 4.

Example 2

(No-WCF)L2  los impuestos para que el gobierno pague  (Session 3)

‘the taxes so that the government could pay [subj.]’

Example 3

(WCF)L2  Necesitan algo en que puede creer  (Session 10)

‘They need something in what (they) *can [indic.] believe’

Example 4
‘But I do not think that it is [subj.] necessarily fair’

Furthermore, given that nominal clauses were far more recurrent in the written production of the L2 and HL learners in the two groups, this category was further classified according to the types found in their written data, including (a) impersonal structures that express possibility, (b) verbs of doubt, (c) verbs of negation, (d) impersonal structures that express opinion and (e) influence/volition.

The production results showed that the most frequent forms and structures that triggered the use of the subjunctive in both the L2 and HL learners across the two conditions involved nominal clauses with negated main clauses, as in Example 5 and impersonal expressions of opinion, as in Example 6.

**Example 5**

(WCF)L2  
*No pienso que la religión es*  
‘I do not think religion *is [indic.]’

(WCF)HL  
*Esto no significa que da la felicidad*  
‘This does not mean that it *brings [indic.] happiness’

**Example 6**

(No-WCF)HL  
*Es imposible que es verdadero*  
‘It is impossible that it is true’

(No-WCF)L2  
*Es importante que ayuda personas*  
‘It is important that (you) help people’

With respect to the use of epistemic clauses, as mentioned earlier in this chapter, the high rate of negated verbs, as in Example 5 above, was likely due to the stance taken by the L2 writers
when responding to the assertions included in the argumentative prompts. On the other hand, the use of impersonal structures, as in Example 6 above, could be seen as a discursive style that attempts to make the writers’ voice less personal while still expressing their opinion on a particular topic.

In terms of the production of subordinate clauses, the only major difference between the L2 and HL learners was found with adjective clauses, as in Example 7, which were more frequent among the HL learners.

Example 7

(WCF)HL  A la persona con ese seguro que no sea afiliado (Session 3)
‘To the person with that insurance that is not affiliated’

Some authors (e.g., Biber, 1988) have argued that a high rate of use in adjectival clauses in writing could be seen as an indication of formality, which, in the case of the HL learners, might be related to their higher level of written complexity (i.e., higher level of subordination), which is further discussed in one of the subsequent sections dedicated to the CAF measures.

Comparison of Error Production

As research has shown (e.g., Montrul, 2007; 2011) the acquisition of the subjunctive and mood selection represents a formidable challenge for both L2 and HL learners whose L1 is English, and full command of this grammatical feature is many times not achieved even among advanced learners. Some of the areas that have been found to be most problematic for L2 and HL learners include identifying and discriminating the morphological, lexical, semantic, and pragmatic implicatures of indicative and subjunctive morphology in variable contexts. As such, errors that involve the production of the indicative form in contexts that necessitate the use of the subjunctive are frequently found in the writing of L2 and HL learners, as in Example 1.
Example 1

(No-WCF) L2  *Es increíble que un país puede mejorar muchas vidas*  (Session 11)

‘It is incredible that a country *can [indic.] improve many lives’

(No-WCF) HL  *Es muy posible que un hombre que tiene un trabajo*  (Session 3)

‘It is very possible that a man that *has [indic.] a job’

Before comparing the L2 and HL learners across the two conditions in terms of their total number of present subjunctive errors and overall accuracy, for this discussion it is relevant to further examine the factors that influence error production in order to understand the outcomes of the treatment on this particular form.

First, in terms of morphology, as Collentine (2013) has argued, while there are some irregular forms (e.g., *tenga* ‘should/may have’), the perceptual saliency of the present subjunctive is probably very low because the thematic vowels that distinguish the indicative and the subjunctive can be ambiguous to learners both functionally and semantically, as shown in Example 2.

Example 2

(WCF) L2  *Es necesario que todos en el mundo ayudan*  (Session 1)

‘It is necessary that everyone in the world *helps [indic.]’

As seen in Example 2, the only morphological difference between the non target-like indicative form *ayudan* ‘they help’ and the target subjunctive form *ayuden* ‘they help’ is a switch between the thematic vowels ‘a’ in the indicative to ‘e’ for the subjunctive. In addition to low perceptual features, some authors (e.g., Collentine, 2010; VanPatten, 1996, 2002) have argued that those grammatical forms with low communicative value, namely the overall meaning that is brought to sentence comprehension, are the least likely to be processed and may never get
acquired. In the case of the subjunctive, the modality conveyed in the subordinate clause through the subjunctive mood, as Collentine (2010) noted, is mostly redundant because the same information is already included in the main clause, whether in the form of nouns (i.e., adjectival clauses), conjunction of sentences (i.e., adverbial clauses) or verbs (i.e., nominal clause), as in Example 3.

**Example 3**

(No-WCF)L2  *Es posible que causan la separación*  (Session 10)

‘It is possible [modality] that they *cause [indic.] the separation’

In this example, the expression *es posible* que ‘it is possible that’ in the main clause elicits the subjunctive in the subordinate clause. In nominal clauses, the subjunctive mood in Spanish, as with *causen* ‘they cause’ [subj.], often expresses the concept of *irrealis*, which entails that something may or may not happen. In the Example 3 above, the information in the subjunctive is redundant, and therefore has very low communicative value, given that the expression of possibility is already conveyed in the main clause, *es posible* que ‘it is possible that’, and the meaning can be simply obtained through the stem of the verb ‘caus-’.

Additionally, the acquisition of the subjunctive with L2 and HL learners also involves understanding pragmatic and abstract concepts, including the distinction between *realis-irrealis*, mentioned earlier, events and states of mind. For example, if a learner is not aware that the subjunctive denotes propositions about events that are not anchored in the actual world, then, errors as in Example 4 may occur.

**Example 4**

(No-WCF)L2  *Porque requiere que hagamos cosas*  (Session 6)

‘Because it requires that we *do [indic.] things’
Considering all of these factors that affect the acquisition and production of this grammatical category, it is not surprising that the accurate production of the present subjunctive was the lowest of all of the forms examined in this study. However, while both the L2 and HL learners in this study made errors, there are some important differences between them that need to be discussed.

First, the comparison between the L2 and HL learners, in terms of the total number of present subjunctive errors produced across all sessions, showed that regardless of the group condition, the HL learners produced significantly fewer errors than their L2 counterparts. In terms of overall accuracy, the L2 learners in the WCF showed 17.5% accurate use across all categories that triggered the use of the present subjunctive, compared with 18.3% for the L2 learners in the No-WCF group. The HL learners in the WCF and the No-WCF groups were considerably more accurate with 62.5% and 59%, respectively, of correct use of the present subjunctive across all sessions. These findings with timed written production are in line with what Montrul (2011) found in spontaneous oral production, namely, that adult Spanish heritage speakers are more accurate than adult L2 learners under communicative pressure.

The fact that the HL learners were overall more accurate than their L2 counterparts begs the question of whether early exposure to Spanish presented an advantage to HL learners over L2 learners. According to Carreira and Potowski (2011), age of acquisition does not provide an advantage because the subjunctive is acquired later at the time when HL speakers’ exposure diminishes due to increased contact with English during schooling. Furthermore, the reduction in input at a time when the subjunctive mood is being acquired could explain, as Carreira and Potowski (2011) noted, why some of the HL learners in this study showed very limited or no
control at all, with the production of the present subjunctive in some of the categories, such as in adjectival clauses, as in Example 5.

**Example 5**

(WCF)HL *Para hacer cualquier cosa que te hace feliz* (Session 7)

‘To do anything that *makes [indic.] you happy’

A closer examination of the specific categories that elicited the production of the present subjunctive also revealed some differences between the L2 and HL learners within and across the WCF and No-WCF groups that are worth discussing. First, previous research with oral production tasks (e.g., Collentine, 1995; Geeslin & Gudmestad, 2008) found that intermediate L2 learners of Spanish were most successful with the production of the subjunctive when the matrix verb expressed volition; however, in this study the L2 learners never produced the present subjunctive when it was triggered by verbs of volition, as in Example 6.

**Example 6**

(WCF)L2 *Los estereotipos causan que personas piensan* (Session 2)

‘Stereotypes cause that people *think [indic.]’

(No-WCF)L2 *Dejar que la gente elige su propia manera* (Session 3)

‘To let that the people *choose [indic.] their own way’

The L2 learners in both the WCF group and No-WCF group were most accurate with the use of the subjunctive in adverbial clauses, followed by impersonal structures that expressed opinion and negated epistemic clauses. The comparison between the L2 and HL across the two conditions showed that the HL learners were more accurate than their L2 counterparts in all of the categories that triggered the use of the subjunctive, except with negated clauses. While this finding might be surprising, a possible explanation for the lower accuracy rate with the use of the
present subjunctive after negated epistemic clauses might be related to the variety of Spanish that
the HL learners acquired at home. Some authors (e.g., Silva-Corvalán, 1994; Torreblanca, 1997)
have found that in some varieties of US Spanish and Mexican-American dialects there is an
increased use of the indicative mood in contexts where the subjunctive would normally occur in
“standard” Spanish, especially with verbs of doubt and in negation, as in Example 7.

**Example 7**

(WCF)HL  *Yo no creo que la globalizacion causa los problemas economicos* (Session 9)

‘I do not believe that globalization causes [indic.] the economic problems’

(No-WCF)HL  *Yo no creo que es justo hacer*  

(Session 2)

‘I do not believe that is [indic.] fair to do’

Considering that most of the HL learners in this study were of Mexican background, it is
possible that what appeared to be an error or deviation of the norm was indeed the result of a
variety of Spanish that has extended the use of the present indicative to other contexts. This
question is particularly relevant when giving feedback to heritage speakers, especially on forms
such as the subjunctive, which are subject to dialectal variation, given that the provision of
grammar correction on forms that are not “standard”, but accepted and used in their community,
could be confusing for them and even affect the perception of their home variety.

Another important aspect of the production of the subjunctive relates to the errors of
“oversupplyance”, that is, the use of the present subjunctive in contexts where the indicative is
obligatory. As noted earlier, in complex sentences learners need to select between the indicative
and subjunctive moods based on morphological, semantic, lexical and pragmatic variables. As a
result of explicit instruction and form focus, previous research (Collentine et al., 2002) has
reported that learners may overextend the use of the present subjunctive to contexts where the
indicative should be used. This study found that both L2 and HL learners produced errors of
oversuppliance of the present subjunctive, as in Example 8.

Example 8

(WCF)L2 Mientras creo que la sanidad sea muy importante (Session 3)
‘While I believe that heath care *is [subj.] very important’

(No-WCF)HL No se si haya problemas a cause de ella (Session 9)
‘I do not now if there *are [subj.] as a cause of it’

A comparison in terms of errors of oversuppliance of the present subjunctive between the
L2 and HL learners across the two conditions showed that they were far more frequent in the
written production of the L2 learners in the WCF group. Additionally, the percentage of L2
learners who overextended the use of the subjunctive was significantly higher in the WCF group
(75%) than in the No-WCF group (25%). The HL learners in the WCF group, by contrast, did
not make any errors of oversuppliance of the subjunctive, whereas 33% of the HL learners in the
No-WCF produced at least one error of this kind. Hence, the combined results in terms of error
production suggest that the treatment did not help L2 learners produce fewer errors with the
present subjunctive; however, there were cases in which the WCF could have induced them to
use it in the wrong contexts, as in Example 9.

Example 9

(WCF)L2 No pienso que el dinero da (de) la felicidad real (Session 6)
I do not think that money *gives [indic.] (gives) [subj.] real happiness
‘I do not think money brings real happiness’

(WCF)L2 Pienso que exista un problema con (Session 7)
‘I think that there *is [subj.] a problem with’
In Example 9, the L2 learner received indication that the indicative form *da* ‘gives’ was incorrect after the negated epistemic clause *no pienso que* ‘I do not think that’. After correcting the error, the same L2 learner, in a subsequent text, produced the present subjunctive *exista* ‘exists’ in a context where the use of the indicative form *existe* ‘exists’ was obligatory. Given that the error was produced with the affirmative epistemic clause *Pienso que* ‘I think that’, it is possible that the learner interpreted the feedback incorrectly and extended the use of the subjunctive to “all” epistemic clauses, both negative and affirmative.

A possible explanation for errors of oversuppliance is that, as Montrul and Perpiñán (2011) noted, mood selection cannot just be accessed by simply observing contrasts between the forms and the context, but instead it must be constructed from complex pragmatic inferences. As such, the limited information that the learners are able to obtain from the WCF could lead to incorrect hypothesis testing and inferences about the use and meanings of the subjunctive that may result in errors of overextension, as in Example 9 above.

**Error Revision**

Similar to what it was found with the previous linguistic forms examined in this study, the L2 and HL learners in the WCF group were able to revise more errors than their L2 and HL counterparts in the No-WCF group, as in Example 10.

**Example 10**

(WCF)S19  *Es mucho más probable que vas (vayas) a encontrar*  (Session 6)

‘It is much more likely that you *are [indic.] (find-subj.) going to find*

(No-WCF)S5  *Es posible que hay una causa*  (Session 9)

‘It is possible that *there is [indic.] a cause’
In Example 10, an L2 learner in the WCF group that received WCF for the non-target like present indicative form vas ‘you go’ provided the correctly revised present subjunctive form vayás ‘go’ during the revision session, whereas the non-target indicative form hay ‘there is’ \[indic.\] produced by the L2 learner in the No-WCF group was not corrected during the revision session.

While these findings provide further evidence of the efficacy of the WCF with error revision, the fact that the proportion of the present subjunctive errors revised by the L2 learners was much higher compared to the HL in the WCF group begs further discussion. As argued earlier, there are certain uses of the present indicative (e.g., in subordinate clauses after negated epistemic clauses) that are accepted in some dialects of Spanish (e.g., Mexican) and varieties of US Spanish. As such, it could be that some HL learners did not perceive that the verb form that was marked in their text was erroneous because that use of the present subjunctive in that context was part of their grammars. In the same vein, it is possible that some HL learners, especially the more advanced ones, decided to ignore some errors because they contradicted their already established mental representations. Another account is that the HL learners simply decided to ignore certain corrections, including present subjunctive errors, in order to focus on the revision of other types of errors (e.g., spelling, content, etc.).

Additionally, it should be noted that the higher percentage of error revision in the L2 learners could be due to differences between the two types of learners in terms of metalinguistic knowledge. Previous experimental research (e.g., Bowles, 2010; Montrul & Perpiñán, 2011) that has compared the knowledge of English-speaking L2 learners of Spanish and HL learners in different tasks and grammatical categories, including tense-aspect and mood, has pointed that HL learners perform better on tasks that tap implicit, intuitive knowledge, whereas L2 learners tend
to do better on tasks that rely on declarative and metalinguistic knowledge. Hence, if according to Montrul and Perpiñán (2011), HL learners know how to use the language but may lack the necessary metalinguistic command of the labels for different language structures, then it is not surprising that, in a task like written error revision that maximizes metalinguistic knowledge, the L2 learners outperformed the HL learners.

Production of the Present Subjunctive forms in New Texts

With respect to the effect of error correction and WCF on the accurate production of specific forms of the present subjunctive that were traceable over time, the results provided some interesting information that is worth discussing. First, it was found that in spite of receiving WCF and successfully self-revising their errors, when the L2 learners produced the same structure that elicited the use of the subjunctive in a new text, the same error type (i.e., present indicative in a context where the subjunctive is required) was made, as in Example 11 below.

Example 11

(L2)S4 Y por eso no pienso que son (sean) justos (Session 2)
‘And because of that I do not think they *are [indic.] (are-subj.) fair’

(L2)S4 No pienso que el dinero da la felicidad (Session 6)
I do not think that money *gives [indic.] happiness
‘I do not think money brings happiness’

In the example above, an L2 learners in the WCF group received indication that the verb form son ‘are’ in the indicative was incorrect in the output and provided the target form sean ‘are’ in the subjunctive during the revision session. Later in session 6, the L2 learner produced the same epistemic negated clause that elicited the use present subjunctive in the subordinate clause, and the indicative form was used again. In the case of the HL learners, there were only
two traceable forms, which do not allow for inferences or conclusions regarding the effect of the WCF with this population.

Considering that the underlying assumption for giving feedback is that it will help learners notice their errors and, subsequently, produce the correct forms, the available data with L2 learners indicate that the WCF treatment was ineffective in helping the learners produce the present subjunctive in new texts. Regarding the ‘no revision’ condition, regardless of the group condition, it was found that the errors that were not revised by L2 and HL learners in one text were also produced again in subsequent texts, as shown in Example 12.

Example 12

(No-WCF)S19 *No pienso que estereotipos son justos*  (Session 2)

‘I do not think that stereotypes *are [indic.] fair’

(No-WCF)S19 *No creo que hay otro país*  (Session 7)

‘I do not believe that *there is [indic.] another country’

(No-WCF)S19 *No creo que la religión es*  (Session 10)

‘I do not believe that religion *is [indic.]’

The excerpt in Example 12 illustrates the case of a learner in the No-WCF group that consistently produced the indicative forms after the negated epistemic clauses *no pienso que* ‘I do not think that’ and *no creo que* ‘I do not believe that’ which elicit the use of the subjunctive. These findings were expected because although written production within the writing-to-learn perspective (Harklau, 2002; Manchón, 2009, 2011) is seen as a vehicle for L2 learning, its potential role in language acquisition is related to the consolidation of linguistic knowledge of grammatical features that are already part of the learners’ declarative knowledge (e.g., William, 2012). Hence, considering that most L2 learners seemed to be at an early stage in the acquisition
of the subjunctive, given their low accuracy and variable morphology, it is not surprising that the same errors were produced again in subsequent pieces of writing and that knowledge restructuring or consolidation though written output was unattainable. This same argument can be applied to the role of feedback, namely, that it can only promote the acquisition of forms when it is provided at the right stage of the learner’s development. In the case of the present subjunctive, as argued earlier in this chapter, the subjunctive mood cannot be learned by simply drawing the learners’ attention to verb morphology, given that its use is constructed from complex pragmatic and semantic inferences. Hence, if the learners do not meet the necessary L2 developmental conditions for cognitive comparison that can enable them to notice the gap, including, conscious awareness of the subjunctive rules, then the feedback may not be just ineffective but may also negatively affect other grammatical categories as shown by the cases of ‘oversuppliance’ in which the learners, after receiving WCF, produced the present subjunctive form in contexts where the indicative was required.

5.4. Complexity and Fluency

The third research question explored the influence of WCF on both complexity, including verbal density and lexical diversity, as well as on fluency, in order to test Truscott’s (1996; 2007) claims that WCF may have a negative effects on other areas of writing development. His assumptions, as outlined in chapter 2, were primarily based on the trade-off hypothesis, which predicts that due to limited capacity, there is tension between complexity, accuracy and fluency, and L2 writers or speakers often focus, either consciously or subconsciously, on one of the three dimensions to the detriment of the other two. To answer this question, the writing of the learners in the WCF and No-WCF groups were compared in terms of complexity and fluency before and after the feedback intervention following a pre-posttest design.
5.4.1. Complexity

With respect to the first measure of complexity, lexical diversity, the results showed no variation between the first and the last texts when comparing the WCF and No-WCF groups, which provides indication that the treatment did not have an effect on the size and lexical variation of the participants.

The second measure of complexity, verbal density, which compared the average number of verbs per t-unit before and after the treatment showed that the two groups experienced a similar decrease with this rate. One explanation for this variation is that the writing of each text was elicited by two different prompts. As such, is possible that the prompt used for the first text (‘do you think that it is important to help other people?’) could have stimulated the use of more complex sentences as compared to the one used in the last session (‘do you think that there is a problem with immigration in the US?’), as in Example 1 with the use of the conditional to express possible scenarios.

(WCF)S4 Si alguien necesita ayuda en algo pequeño, (Session 1)
como su tarea, es muy fácil ayudarle.

‘If someone needs help in something small, like his/her homework, it is very easy to help him/her’

(WCF)S4 Esta es buena para la diversidad del país, (Session 11)
pero crea un problema con los trabajos disponibles

‘This is good for the diversity of the country, but it creates a problem with the available jobs’

In addition to verbal density, as a measure of syntactic complexity, it should be noted that the two groups were also comparable in terms of the total production and occurrence of the
linguistic forms examined in this study. This is particularly relevant in the case of the structures that elicited the use of the present subjunctive, which can be seen as an indication that the treatment did not have a negative effect on the production of complex sentences and subordinate clauses.

The overall results are in line with Van Beuningen et al. (2012) who found that comprehensive WCF did not have a negative effect on any of the two measures of complexity, and also contradicts Truscott’s claims (1996, 1999, 2007) as error correction did not motivate the students to use ‘simplified’ writing by avoiding the forms and structures that had been corrected.

Regarding error avoidance and complexity, a survey that the students answered at the end of the data collection period offered some interesting information that should be considered. When the learners were prompted to indicate how often they tried to write less complex sentences or ideas to avoid grammar errors, most of the participants in the WCF group (45%) and the No-WCF group (38%) indicated that they did it ‘occasionally’, followed by ‘rarely’, WCF group (30%) and No-WCF (28.6%) and ‘very rarely’, WCF group (10%) and No-WCF (19.1%). There are two important implications that emerge from these results. First, they provide evidence from two differences sources, written data and learners’ perceptions, that indicate that the treatment did not have an impact on error avoidance. Secondly, there were learners in the control group, who did not receive feedback or explicit attention to form, but decided to simplify their writing to avoid errors. This finding is relevant because it highlights the fact that learners may intentionally decide to center their attention on certain areas of their L2 writing (i.e., accuracy, complexity, etc.), instead of others. As such, a possible consequence of this is that the learner’s conscious attention to a particular CAF component may influence the outcome of their development, regardless of the goal of the intervention.
5.4.2. Fluency

Based on previous WCF research (Hartshorn, 2008; Hartshorn, et al., 2010; Vyatkina, 2010) fluency was operationalized as the total number of words that a student was able to include in their writing within a particular period of time. It was hypothesized, following the Trade-off hypothesis (Skehan, 1998) and the Monitor hypothesis (Krashen, 1982), that the corrective feedback could inhibit fluency if the WCF caused the learners to monitor their production more carefully. Additionally, based on the Affective Filter Hypothesis (Krashen, 1982), error correction could cause higher anxiety and raise the learners’ affective filter, and thus hinder fluency in the L2.

The comparison between the participants in the WCF and No-WCF conditions before and after the intervention showed no differences in terms of fluency (i.e., total number of words per minute), which suggests that the treatment did not interfere this area of writing development. In contrast to Hartshorn et al. (2010), who also compared the effects of indirect comprehensive WCF on fluency and found that the treatment group was slightly less fluent than the control group in the posttest; the results of this study offer the opposite picture, as the participants in the WCF group increased overall their fluency over time, outperforming the No-WCF group. Several explanations may account for these differences. First, while both treatments involved short timed writing with frequent feedback and error revision, the methodology in Hartshorn et al. (2010) and other studies that tested their “dynamic” WCF (Evans et al., 2010; Evans, Hartshorn & Strong-Krause, 2010) were significantly more time consuming and required far more focus on form than the one employed in this study. Considering that the treatment in Hartshorn et al. (2010) entailed multiple draft revisions, rewriting of texts and classroom discussions centered on the most frequent types of errors produced by the students, it is not surprising that such emphasis
on form led to some decline in fluency given the strong focus on accuracy. However, Hartshorn et al. (2010) and many previous studies have examined the effects of WCF on accuracy development without looking at the perceived impact of the treatment. Thus, it is argued in this study that information solely based on the learners’ written data provides a rather incomplete picture of the intervention as it does not take into account the participants’ view of their own development. In this study, a look at the learners’ perception of their writing development, as shown by the survey results, revealed that the quantitative results on fluency were in line with the perception of the learners in the WCF as they all reported the ability to write faster at the end of the treatment.

Furthermore, the fact that the learners in the No-WCF group experienced a small decline in fluency in the pre-posttest could be due to various reasons. It is possible that the small decay in fluency was simply part of fluctuation, as observed with the accuracy rate of some of the forms in this study, which could be due to factors such as knowledge of the topic, the instruction that they were receiving, or their differential value of the CAF components. In this sense, the differences between the two groups in the amount of attention paid to grammar could explain the variation in the fluency results. The perception results revealed that the participants in the No-WCF group paid considerably more attention to grammar (I always pay attention = 9.5%; very frequently = 76.2%; occasionally= 9.5%) than their counterparts in the WCF group (I always pay attention = 5%; very frequently = 45%; occasionally = 45%) when writing their texts. The fact that the participants who received CF reported less attention to grammar during writing than those who did not receive feedback begs further consideration. A possible explanation is that the learners in the WCF group may have decided, consciously or subconsciously, to focus less on grammar because they knew that were receiving feedback for each of their writings, while the
participants in the no feedback condition were fully responsible for monitoring their accurate production. These findings bring a new question that has not yet been considered in CF research, namely, that researchers and instructors should not be only concerned with the prospect that learners, due to grammar feedback, may put excessive emphasis on accuracy, but also consider the opposite, that is, that constant WCF could actually result in lower attention to grammar and higher dependency on the teacher’s feedback.

In conclusion, the combined findings of complexity and fluency suggest that WCF did not have a harmful effect on the complexity and fluency measures employed in this study. Considering the SLA theory, the comparison between the two groups indicate that, in general terms, error correction did not lead to greater Monitor use (Krashen, 1982), which involves more conscious L2 knowledge resulting in a potential decrease of fluency. Likewise, the results suggest that error correction did not have a negative effect on the learners’ Affective Filter (Krashen, 1982) as the comparison between the two conditions revealed that the WCF group did not try to avoid difficult constructions or focus more on form and less on meaning.

5.5. Implications

The findings of this study provide theoretical and practical implications for both language practitioners and researchers related to the role of L2 written production and revision in the accuracy development of specific linguistic L2 forms in a foreign language context in addition to further considerations regarding research design for future WCF studies.

The first implication of this study relates to the role of indirect comprehensive WCF feedback as a self-editing tool to assist L2 learners with text revision. Language instructors, especially in composition courses, are concerned with ways to help their students improve the accuracy of an initial text or draft, and they tend to spend a lot of time and energy on WCF. As
such, some authors (e.g., Bitchener & Ferris, 2012; Ferris, 2010) have argued for the importance of identifying strategies to help students more successfully revise and edit their work. In this sense, the findings of this study provide some important implications both regarding the type of WCF treatment implemented and the participants’ ability to self-correct their errors during revision sessions. First, the overall revision results indicate that the indirect WCF, in which the students’ errors were marked and then the students had to provide the correct target form, was largely effective in helping improve the accuracy of the learners’ initial texts. The students who received feedback were not only able to revise a larger number of errors as compared to the No-WCF group, but were also more accurate with their corrections, as they were able to provide over 70% correct revisions for all twelve linguistic categories that were analyzed. It should also be noted that the students in the WCF group were more accurate when revising ‘treatable’ errors (e.g., Ferris, 1999), which required the application of systematic rules, including canonical gender marking, verb tense, subject verb agreement or articles. Error revision was less accurate, in word choice and spelling, which are considered ‘untreatable’, that is, forms that are not rule-governed, as well as in those linguistic features regarded as more advanced or complex, such as subjunctive mood, which requires longer periods in order to be acquired. Still, it should be noted that it is very likely that if the learners had not received explicit indication at the end of their texts in the form metalinguistic comments about the error type, their ability to notice and correct their errors would have been significantly lower. A pilot study conducted to evaluate different feedback strategies found, based on their comments and percentage of correct revisions, that the students had problems noticing and revising some of their errors, especially those forms such as the subjunctive that were not part of their developing grammar or minimally acquired, by means of only indirect WCF (e.g., underlined errors). Moreover, it is also important to keep in mind
when assessing the efficacy of the treatment in error revision, that the participants in this study were either minoring or majoring in their L2 and, as such, their metalinguistic and explicit knowledge on grammatical rules and categories was higher than that of most FL learners.

Another important point to consider is the learners’ behavior when correcting their errors during the revision sessions. The results of the end-of-treatment questionnaire revealed that when prompted to indicate whether they asked their instructor for clarification or help during the revision sessions (ranging from 1 -‘always’ to 6 -‘never’), the majority of the students indicated that they did it ‘rarely’/’very rarely’, WCF group (M=4.4; SD=1.54) and No-WCF group (M=5.06; SD=1.2). These findings suggest a high degree of confidence in terms of error revision with the resources available to them and that they did not need their instructor for further assistance.

In addition to the learners’ ability to self-correct their errors with WCF, the findings of this investigation offer valuable information regarding the type of errors that learners are able to revise without feedback help. The vast majority of the errors corrected by the participants in the No-WCF group, including both L2 and HL learners, corresponded to accent marks, spelling and word choice, which underline a focus on mechanical and surface errors and also their inability to notice grammatical errors in their output. As such, these results point to the fact that some sort of form-focus intervention, such as that of WCF, is necessary for learners to be aware of and correct their errors in order to improve the accuracy of an initial text by means of revision.

Lastly, it should be noted that the comparison between L2 and HL learners pointed that the issue of correction and error revision can be more complex when teaching HL learners. As indicated in the discussion, instructors need to be aware that there are features in the HL students’ production that are the result of bilingual acquisition, incomplete acquisition or contact
with English that may affect the extent to which they can notice and respond to error correction. Likewise, we hypothesized that unlike L2 learners; HL learners, due to their more established grammars, may reject or ignore some of the corrections they receive, and as such this population might be less amenable to CF treatments.

The second implication relates to the effects of revision with and without WCF on the accurate production of specific linguistic forms, which is the central question of the grammar correction debate and one that has generated great amount of interest among SLA researchers and classroom teachers. Before discussing the findings and implications of the results, it is important to note that the purpose of this study was not to determine the best way to provide WCF, but to test the extent to which grammar revision and WCF, which is the type of feedback most commonly used by language instructors, can lead to the students’ subsequent accurate production in a variety of linguistic forms. Still, from a SLA perspective, the treatment was implemented following some of the principles stated in previous research, including that feedback must be meaningful, timely and constant, and manageable for the student (Evans et al., 2010; Evans, Hartshorn & Strong-Krause, 2010; Hartshorn et al., 2010). In order to make the WCF meaningful, the feedback was contextualized, individualized and embedded in the discourse generated by the learner, and included time for self-editing. Additionally, the WCF was timely and constant as the students produced and received feedback at the beginning of each class period during the treatment period. Finally, the feedback appeared to be manageable for the students as the majority of the grammar corrections were revised, which suggests that they had the time and ability to process them.

Furthermore, the forms that were analyzed in this study, namely, canonical gender marking, non-canonical gender marking, article omission and the present subjunctive can be seen
as an excellent measure to test the efficacy of WCF at different levels as these forms vary in how they are acquired (i.e., rule-based grammatical features, lexically stored), their degree of complexity, and the level of acquisition ranging from minimally to highly acquired forms. The findings of this study provided relevant information related to the effect of the WCF on the accuracy development of each linguistic form. First, when comparing the two conditions at the group level, the provision of WCF on highly acquired forms, as shown by their accuracy rate, (i.e., canonical gender marking), did not make a difference in terms of accuracy development. In the same vein, the results at the individual level showed that the lack of feedback did not have a negative effect on accuracy as the majority of the students either increased their accurate production or it remained stable over time.

From a pedagogical perspective, the fact that there were many learners who, without any feedback, were able to improve their accuracy on partially acquired forms (i.e., canonical gender marking, definite articles) underscores the potential of pushed output in L2 acquisition. Hence, as some authors have argued (e.g., DeKeyser, 2007; Housen & Kuiken, 2009; Manchon, 2010; Swain, 1998; 2000), learners should be given ample opportunities for output practice, given that the process of repeated retrieval and deployment of knowledge that result from it may create the necessary conditions (i.e., hypothesis testing, self-monitoring, etc.) that can facilitate the consolidation of existing knowledge. In addition to the role of output, the impact of the input in written production should born in mind. Since the learners had access to the prompts in the L2, as it is the common procedure in most language classrooms, it was argued that the students had time to read and pay attention to the forms and structures in that input (i.e., nouns, articles, verb forms) and potentially incorporate them into their output. As such, the accurate use of these forms could have been the result of employing the information in the input (i.e., prompts) rather
than a product of their own grammatical knowledge. This is a question that has not been considered in previous WCF research, which both researchers and teachers need to keep in mind since it is self-evident that any written input that the learners receive and respond to is likely to influence what they eventually produce.

With respect to the effect of the WCF on the accurate production of complex linguistic features, the findings related to the present subjunctive pointed that the treatment was ineffective in helping the feedback group improve their accuracy over time. It is possible that, as Storch and Wigglesworth (2010) suggested, indirect feedback cannot help increase the control of a linguistic form that is not partially internalized or is minimally controlled, as it was the case for the present subjunctive. However, assessing the potential effect of WCF solely based on different types (i.e., indirect vs. direct WCF) can be risky because it does not take into account what the learners do with the feedback or the type of L2 form or structure for which the feedback is given. For instance, in the case of the present subjunctive and non-canonical gender forms in this study, the learners who received feedback were generally able to notice and self-correct their subjunctive errors; however even after revising their errors, the learners continued to make the same errors again. If we attribute incorrect use of an L2 form to the type of WCF provided, we would simply be ignoring other factors, such as the complexity of the linguistic forms or the intricate process of acquisition, that play a more crucial role in determining the extent to which learners are able to produce an L2 form correctly in subsequent production.

An important theoretical question that emerges is whether the type of noticing that results from the WCF can lead to accuracy development. According to Schmidt (2001) “SLA is largely driven by what learners pay attention to and notice in target language input and what they understand the significance of noticed input to be” (p. 3-4). In this study, based on the
indications of previous research (e.g., Bitchener, 2012; Bruton, 2009a; Chandler, 2003; Sachs & Polio, 2007; Storch, 2010), all of the students were given the opportunity to revise their errors in order to not only ensure that the corrections were attended to, but also because it is hypothesized that error revision may contribute positively to learning. Then, ‘noticing’ in the context of this study is to what Bitchener (2012) refers to as “the process of registering that there is a mismatch or gap between a learner’s interlanguage output and the target L2 input” (p. 351). With this in mind, the results, particularly in the case of the present subjunctive and non-canonical nouns, showed that error noticing, in the form of revision, did not result in the uptake of the same forms over time. Thus, an important implication of these findings is that successful error revision should not be seen as evidence of learning or as an indication of future accurate production.

Even though some proponents of grammar feedback may argue that yet longer interventions with even more frequent comprehensive WCF may bring about some changes in accuracy, it is important to remember that the treatment implemented in this study offered the students far more opportunities for revision and feedback (i.e., a total of nine sessions), and was more intense (i.e., the WCF was provided three times per week) than what can be found in most language classrooms. A practical consideration for the FL classroom is that the time and effort that instructors invest in providing comprehensive and individualized WCF will most likely have no effect when looking at the accuracy development of the forms that they corrected for their students. A second question that arises for language teachers and curriculum developers is the goal of giving learners, especially in FL settings, intense and customized WCF in the current context of communicative language teaching. Teachers may still use comprehensive WCF to help learners improve the accuracy of a previous draft; however the results of this study in the view of SLA theory offer compelling evidence to understand that this approach to grammar
correction, as it is typically implemented, will not make a difference for most L2 learners in terms of their acquisition of L2 forms.

Returning to the initial debate of the role of grammar feedback, and more specifically to the question of whether WCF can help increase the accurate use of L2 forms, the findings of this study lend support to Truscott’s claim (1996; 2004; 2007) and contradict the claims of previous research (e.g., Evans et al., 2010; Hartshorn et al., 2010; Van Beuningen, 2012) that comprehensive WCF can lead to improved accuracy over time. An explanation for these opposing findings can be found in the study design.

We believe that a significant contribution of this study relates to the type of design employed to address one of the main limitations of previous research, which is the lack of qualitative analyses of learners’ accuracy performance on particular L2 forms and structures. Unlike the current study, research on comprehensive WCF has typically assessed improvements in accuracy by means of broad measures that involve ratio/percentage of correct words, clauses (e.g., Evans et al., 2010; Hartshorn et al., 2010; Truscott & Hsu, 2008), or categories such as ‘grammatical’ and ‘non-grammatical’ (Van Beuningen, 2012) or ‘mechanical, lexical and grammatical’ errors (Hartshorn & Evans, 2012). While these studies offer some valuable information, they do not provide actual evidence that the knowledge obtained after receiving feedback or revising a specific L2 form in one piece of writing (i.e., pretest) is then applied to a second (posttest) or third text (i.e., delayed posttest). Also, as some authors have observed (e.g., Bruton, 2009a; Storch & Wigglesworth, 2008), when looking at the errors made by a particular learner, there are many instances in which the errors from one text bear no relation with the errors made in a subsequent text, and, as such, combining different types is likely to provide a distorted picture of the learning and SLA effects of any WCF treatment. Furthermore, if our goal
is to conduct WCF studies that test the efficacy of grammar correction in terms of SLA, assessing accuracy development by means of categories that contain different error types is counterintuitive, as it does not take into account the premise that each grammatical form is acquired independently. Thus, in order to help inform SLA, research on WCF should not only compare the number or percentage of errors between groups, but also explain how specific L2 forms are acquired and explain why, in light of the results, the WCF did (not) facilitate the acquisition of L2 forms.

5.6. Limitations and Directions for Future Research

Although the findings of this study provide evidence of the inefficacy of comprehensive WCF for the accurate production of specific linguistic forms, namely canonical gender marking, non-canonical gender marking, definite articles in obligatory contexts and the present subjunctive in L2 and HL learners of Spanish, there are a number of limitations that need to be acknowledged and considered for future research.

Although the four grammatical categories analyzed in this study were carefully chosen to provide a comprehensive understanding of the efficacy of the feedback at various levels, future research should continue investigating other specific rule-based and item-based forms to help understand, not only how and under what circumstances a WCF treatment may (or may not) work, but also the role of output practice within the control (i.e., no-feedback) condition.

Differences between the contexts of acquisition need to be considered in future research. In contrast to most of the available research (e.g. Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2009; Bitchener & Knoch, 2010; Ellis et al., 2008; Evans et al., 2011; Evans, Hartshorn & Strong-Krause, 2011; Hartshorn et al., 2010; Sheen, 2007; Sheen, 2010; Truscott & Hsu, 2008; Van Beuningen et al., 2012) that have investigated WCF in L2 or
immersion settings, the present study was implemented in a foreign language classroom. Although this could be considered a strength of the investigation, as FL environments may avoid “the possible adulteration/contamination over time from extraneous influences apart from the corrective feedback” (Bruton, 2009, p. 605), it is clear that there are major differences between the FL and L2 learning contexts, including the type of instruction that the learners receive, the amount of input that they have access to both inside and outside the classroom or their motivation to improve their grammar. As such, it is possible that these factors may have influenced the amount of attention and effort that the learners dedicated to the WCF and ultimately the outcome of the treatment.

With respect to the participants in this study, although the proportion of HL learners is comparatively higher than what we can find in most FL classrooms in the U.S, future studies that aim to compare the effects of WCF with L2 and HL learners should attempt to balance the number of participants across conditions.

A potential limitation of this study relates to the “absence of evidence” principle, especially in the case of the HL learners. Even though the use of meaningful individual production tasks helped preserve the ecological validity of the study and reflects the classroom reality, it could be the case that further evidence of the (in)efficacy of the WCF was missed given that forms or structures that were corrected were not produced again by the same participants in subsequent writing. Hence, as mentioned in the previous chapter, future studies could complement the data from free-writing tasks with tailor-made posttest or experimental tasks purposely designed to assess the learning outcomes of the WCF, particularly, with those forms or structures that tend to appear less frequently in the output (i.e., present subjunctive, non-canonical nouns, etc.).
There are also a variety of factors that may have influenced the amount of attention and effort that the participants in this study devoted to grammar correction and accuracy, including the type of writing task or how the task and accuracy were graded. The type of genre employed in this study, argumentation/exposition, is commonly considered as more cognitively demanding (Shin, 2008). As such, it is possible that less demanding tasks could have contributed to more focus on form and, accordingly, greater accuracy. Furthermore, if we take into account achievement goal theory, the impact of grades or the extent to which accuracy is assessed on written tasks is likely to have influenced the amount of attention and, as Dlaska and Krekeler (2013) have observed, affected the outcome of the feedback. Thus, it is possible that if the learners’ accurate production were graded heavily, their attention to form and correctness could have been greater.

Lastly, we would like to echo Bruton (2009b) and call for more studies in classroom environments, not just for the sake of adding “some ecological validity” to our research, but because WCF is ultimately pedagogical, and, such as, should be contextualized within clear decision-making frameworks and under authentic classroom circumstances. In this regard, future research should also begin to incorporate the voice of the learners if we view feedback a “social act” (Lee, 2008) where the WCF is not invariable, but rather interacts with other internal and external factors.

5.7. Conclusions

Contrary to previous research (Evans et al., 2010; Evans, Hartshorn & Strong-Krause, 2010; Hartshorn et al., 2010; Van Beuningen et al., 2012), this study provides evidence that comprehensive WCF was largely ineffective in helping L2 and HL learners improve their accurate production over time in four distinctive L2 features, namely, canonical gender marking,
non-canonical gender marking, definite articles in obligatory contexts and the present subjunctive. The results show that WCF helped L2 and HL learners revise significantly more grammatical errors than their counterparts in the error revision without WCF condition; however, when comparing the two groups in terms of accuracy development, the data suggest that the attention to and revision of more errors did not lead to higher control of any of the L2 forms, except in the use definite articles – but only at the group level. Additionally, this investigation sheds more light on the effects of WCF on other areas of language development. The results indicate that the feedback treatment did not have a negative impact in the measures of written complexity and fluency employed in this study, which adds to what other studies have found (Evans et al., 2010; Evans, Hartshorn & Strong-Krause, 2010; Hartshorn et al., 2010; Van Beuningen et al., 2012).

Whereas previous research has only conjectured the extent to which errors could be responsive to comprehensive WCF, this investigation offered empirical evidence by tracing its effect over time on a selection of grammatical forms that both L2 and HL learners produced in free-written production in a real classroom environment. Given that this is the first study to examine the accuracy development and retention of L2 forms, further research is clearly needed to contribute to the creation of a larger research body in order to accumulate evidence on the effects of both WCF and written production on the accuracy development of different L2 forms and structures so that both teachers and researchers can be informed on the best pedagogical practices in the language classroom.
REFERENCES


Bitchener, J. (2012). The language learning potential of written corrective feedback. *Journal of*


Bruton, A. (2009b). Improving accuracy is not the only reason for writing, and even if it were…. *System, 37* (4), 600-613.


Carreira, Maria. (2007). Teaching Spanish in the U.S.: Beyond the one-size-fits-all paradigm. In


Ferris, D. R. (2010). Second Language Writing Research and Written Corrective Feedback in


(Eds.), *The Acquisition of French in Different Contexts: Focus on Functional Categories* (pp. 175-205). Amsterdam, Netherlands: Benjamins.


Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the
305-313.

Press.


Journal, 66(2), 140.

Biometrics, 33, 159-174.

Larsen-Freeman, D. (2006). The Emergence of Complexity, Fluency, and Accuracy in the Oral
and Written Production of Five Chinese Learners of English. Applied Linguistics, 27(4),
590-619.

Journal, 22, 1-16.

Journal of Second Language Writing 17, 144–164.

Lee, I. (2013). Research into practice: Written corrective feedback. Language Teaching, 46,
108–119.


Philadelphia: John Benjamins.


McLaughlin, B. (1990). “Conscious” versus “unconscious” learning. TESOL Quarterly, 24, 617-


Ullman, M. T. (2001). The Neural Basis of Lexicon and Grammar in First and Second Language:


APPENDIX A: SAMPLE OF THE TREATMENT

A.1. WCF Group (Final version with feedback and revisions)

EDEL 13

(article omitted) (el) dinero es algo vital y importante. Toda piensan que es algo que no lo puede vivir sin. Pero, la pregunta es si (el) dinero es la cosa importante, o las cosas en su vida. Sin dinero, no puede comer, paga cuentas, o hace (vive) mucho. No hay mucho a hacer sin dinero, porque muchas cosas ahora son muy caros. (las) Familias que son pobres luchan cada día con cuentas y comida (comida). Pero en el otro lado, (en opuesto) hay personas que son muy ricos. Tienen más dinero de que puede usar. Estos (estos) personas son feos (feos), además que tienen mucho dinero. Compran coches grandes, ropa y casas. Pero nunca pueden comprar el amor de una familia. Un (una) familia que no tienen dinero no tienen (tiene) muchas cosas materiales, pero tienen (tiene) el amor de su familia.

Buena reflexión sobre el dinero!

Please self-correct your errors. Check gender agreement, article omission, and word choice.

Add your correction in paranthesis next to the form in bold face. You are free to use a dictionary or other online sources to help you rewrite. ¡Gracias!

EDEL 14

(En los) estados unidos, las armas de fuego son un derecho por (para) la gente. En otros países, los ciudadanos (ciudadanos) no pueden tener las armas. Los ciudadanos (ciudadanos) de los estados unidos tienen un derecho que muchos otro no tienen, pero, ahora hay mucho debate sobre las armas fuegos que tiene mucho poder. Cuando están en (las) manos malas, (malas) los (las) armas son muy peligrosas, pero, las armas fuegos deberían (ser) (missing verb) legales. La única cosa que necesita cambiar es la educación de (las) armas fuegas. A todos deberían aprender mucho y toma (toman) clases sobre las armas fuegas (de fuego), porque si alguien sabe mucho, hay un ambiente mas

Muy buen ejercicio!

Please self-correct your errors. Check gender agreement and word choice.

Add your correction in paranthesis next to the form in bold face. You are free to use a dictionary or other online sources to help you rewrite. ¡Gracias!

Notes. Boldface = indirect WCF; In parenthesis = revised errors
APPENDIX A.2. No-WCF group (Final version with revisions)

EDEL 13
Sí creo que el dinero da la felicidad pero depende en la vista. Las personas sin el dinero (dinero) todos sus vidas no pueden hacer muchas cosas que quieren (quieren) hacer. Pero con el dinero debería ser (sería) más feliz porque pueden vivir en el forma que les gustaría. Pero muchas personas que ya el dinero en sus vidas no saben la importancia del dinero. Por ejemplo la gente como, Paris Hilton debería (sería) muy triste sin el dinero porque ella no sabe la vida sin lo.

Buena reflexión sobre el dinero

Read through your text and revise it. Remember to add your correction in parenthesis next to the previous form. You are free to use a dictionary or other online sources to help you revise. Gracias!

EDEL 14
Hay definitamente en (un) problema e (en) los Estados Unidos con las armas de fuego. Quizás el problema es que mucha personas pueden obtener las armas sin los documentos válidos. Creo que deberían ser (sería) legal porque la gente tiene que proteger sus familias especialmente si los criminales tienen las armas con y sin los documentos necesarios. En las noticias hay siempre cuentas de como (cómo) y (un) criminal roban (roba) la gente inocente pero con las armas debería seguro más o menos y tiene la habilidad (habilidad) para defenderse.

Buen uso de los argumentos para apoyar tu posición

Read through your text and revise it. Remember to add your correction in parenthesis next to the previous form. You are free to use a dictionary or other online sources to help you revise. Gracias!

Notes.: In parenthesis = revised errors
APPENDIX B: LANGUAGE BACKGROUND QUESTIONNAIRE

(This information will be kept confidential)

1. Age:_________  2. Gender: ___________  3. E-mail:___________________

2. Place of birth:  City ___________________  Country:_______________________

3. What language/s did you speak at home as a child? ________________________

4. What language does your mother speak? _____________ Father speak? _______

5. In which languages were you educated (i.e., language of instruction
   Elementary School: _____________High School: ___________University:
   __________________

6. Where did you learn Spanish?___________________________________________

7. When did you start studying Spanish? (mark with an X)
   Elementary _____, Jr. High _____, High school _____, University _____

8. Have you ever spent time in a Spanish speaking country? (circle) YES / NO   If so, where?
   _________________________ how long ago?________________ for how long were you
   there?________

9. Are you currently taking any other Spanish courses at UIUC? (circle) YES / NO   If so, which
   one/s?
   __________________________

Rate your current overall language ability in SPANISH
   1 = understand but cannot speak
   2 = understand and can speak with great difficulty
3 = understand and speak but with some difficulty
4 = understand and speak comfortably, with little difficulty
5 = understand and speak fluently like a native speaker

Rate your current overall language ability in ENGLISH
1 = understand but cannot speak
2 = understand and can speak with great difficulty
3 = understand and speak but with some difficulty
4 = understand and speak comfortably, with little difficulty
5 = understand and speak fluently like a native speaker

10. What is your dominant language? (mark with an X) English______, Spanish______, Both ______

11. If you learned or were exposed to Spanish as a child, please complete the second page

1. Where are your parents/caregivers from?
   Mother: __________          Father: ______________

2. At what age did you first begin to learn Spanish?

3. At what age did you first begin to learn English?

4. Did you begin to speak both Spanish and English before age 5? (circle one)
   Yes          No

5. What languages did you hear in your home between the ages of birth-5 years? (circle all those that apply)
   Spanish      English      Mixed      Other (specify) _________

6. What languages did your parents/caregivers use mostly when speaking to you?
   Spanish      English      Mixed      Both      Other
7. What languages did you use mostly when speaking to your parents/caregivers?
   Spanish   English   Mixed   Both   Other

8. Do you have siblings?
   Yes   No   how many?   Are they older or younger?

9. What language/s did you use when speaking with your siblings?
   Spanish   English   Mixed   Both   Other
10. What language/s did your siblings use when speaking with you?
    Spanish   English   Mixed   Both   Other

11. How often did you use Spanish between the ages 6-10?
    always   often   seldom   never

12. Who did you speak Spanish with?

13. How often did you use Spanish between the ages 11-13?
    always   often   seldom   never

14. Who did you speak Spanish with?

15. How often did you use Spanish between the ages 13-17?
    always   often   seldom   never

16. Who did you speak Spanish with?

17. What language(s) do you use at home? How often do you use them? Please indicate percentage:
    English _____ %   Spanish _____ %

18. In what situations do you use Spanish?
APPENDIX C: LIST OF WRITING PROMPTS

1. ¿Crees que es importante ayudar a otras personas? ¿Porqué?
   ‘Do you think it is important to help other people? Why?’

2. ¿Qué es un estereotipo? ¿Piensas que son justos?
   ‘What is a stereotype? Are they fair?’

3. ¿Crees que la sanidad debe ser pública en los Estados Unidos?
   ‘Do you think health care should be public in the US?’

4. ¿Crees que existe igualdad entre los hombres y las mujeres?
   ‘Do you think there is equality between men and women?’

5. ¿Piensas que el mundo sería mejor sin tecnología?
   ‘Do you think the world would be better without technology?’

6. ¿Piensas que el dinero da la felicidad? ¿Por qué sí o por qué no? Usa argumentos y ejemplos para tu posición.
   ‘Do you think money brings happiness? Why? Use arguments and examples for your position’

7. ¿Piensas que existe un problema en los Estados Unidos con las armas de fuego? ¿Crees que deberían ser ilegales? Explica por qué y usa argumentos.
   ‘Do you think there exist a problema in the United States with firearms? Do you think the should be illegal? Explain why and provide arguments.

8. ¿Estás de acuerdo con que la edad legal para beber alcohol deba ser 21? ¿Por qué? ¿Puede ser un problema para los estudiantes universitarios?
   ‘Do you agree that the age to legally drink should be 21? Why? Do you think it is a problem for
9. ¿Qué es la globalización? En tu opinión, ¿crees que hay problemas a causa de la globalización? Explica por qué y usa argumentos.

‘What is globalization? In your opinion, do you think there are problems because of globalization? Explain why and use arguments’

10. ¿Crees que la religión es importante en la sociedad? ¿Las personas necesitan la religión en sus vidas? ¿Existen problemas a causa de la religión? Explica por qué y usa argumentos.

‘Do you think religion is important in society? Do people need religion in their lives? Are there any problems because of religion? Explain why and provide arguments’


‘What is immigration? Do you think it is a problem in the United States? Is it good for society? Explain why’
APPENDIX D: PERCEPTION QUESTIONNAIRE

D.1. Experimental (WCF) Group

1. The feedback on the Ejercicios De Escritura Libre (EDEL) helps me notice my grammar errors.
   - Competely Agree
   - Mostly Agree
   - Slightly Agree
   - Slightly Disagree
   - Mostly Disagree
   - Completely Disagree

2. The feedback on the EDEL makes me more aware of my grammar errors in Spanish.
   - Completely Agree
   - Mostly Agree
   - Slightly Agree
   - Slightly Disagree
   - Mostly Disagree
   - Completely Disagree

3. I know how to correct the grammar errors that my instructor indicated to me.
   - Always
   - Very Frequently
   - Occasionally
   - Rarely
   - Very Rarely
   - Never

4. I am able to correct the errors that were highlighted in boldface for me.
   - Always
   - Very Frequently
   - Occasionally
   - Rarely
5. I ask the instructor when I do not know how to correct a grammar error in my Ejercicios de Escritura Libre (EDEL).

- Very Rarely
- Never

6. I make an effort not to make the same grammar errors when I write the EDEL.

- Always
- Very Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

7. I try to remember the errors I have made before when I am writing the EDEL.

- Always
- Very Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

8. I think it is important to have my errors corrected for me.

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
• Completely Disagree

9. I want to write more accurately in Spanish.
  • Completely Agree
  • Mostly Agree
  • Slightly Agree
  • Slightly Disagree
  • Mostly Disagree
  • Completely Disagree

10. I want to improve my Spanish.
  • Completely Agree
  • Mostly Agree
  • Slightly Agree
  • Slightly Disagree
  • Mostly Disagree
  • Completely Disagree

11. I try to improve my grammar because I want to obtain a good grade in my coursework.
  • Completely Agree
  • Mostly Agree
  • Slightly Agree
  • Slightly Disagree
  • Mostly Disagree
  • Completely Disagree

12. I have learned from the errors I self-corrected after receiving them in boldface.
  • Completely Agree
  • Mostly Agree
  • Slightly Agree
  • Slightly Disagree
  • Mostly Disagree
  • Completely Disagree
13. I have learned new grammar concepts from the corrective feedback I have received.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

14. The grammar correction has helped me obtain better command of grammar concepts that I already knew.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

15. The corrective feedback has helped me learn grammar concepts that I previously did not understand.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

16. I believe the corrective feedback has helped me decrease the frequency of some of the grammar errors I make.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

17. If so, indicate the error types that have been decreased as a result of the grammar corrections received.
18. Please indicate any other type of errors that you think you have reduced:

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

19. I feel anxiety when I see my errors marked.

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

20. I feel overwhelmed with the number of errors that I have to self-correct.

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
21. The time I spent self-correcting my errors is worthwhile given how much I learn from it.

- Completely Disagree
- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

22. I think the amount of grammar correction in the course is enough.

- Completely Disagree
- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

23. I gain knowledge about grammar from reading texts in Spanish.

- Completely Disagree
- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

24. I gain knowledge about grammar from a Spanish grammar book.

- Completely Disagree
- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree
25. I gain knowledge about grammar from other courses I am taking.
   - [ ] Completely Agree
   - [ ] Mostly Agree
   - [ ] Slightly Agree
   - [ ] Slightly Disagree
   - [ ] Mostly Disagree
   - [ ] Completely Disagree

26. I gain knowledge about grammar from interacting with other speakers of Spanish.
   - [ ] Completely Agree
   - [ ] Mostly Agree
   - [ ] Slightly Agree
   - [ ] Slightly Disagree
   - [ ] Mostly Disagree
   - [ ] Completely Disagree

27. I gain knowledge about grammar from online activities.
   - [ ] Completely Agree
   - [ ] Mostly Agree
   - [ ] Slightly Agree
   - [ ] Slightly Disagree
   - [ ] Mostly Disagree
   - [ ] Completely Disagree

28. I can write faster in Spanish than at the beginning of the semester.
   - [ ] Completely Agree
   - [ ] Mostly Agree
   - [ ] Slightly Agree
   - [ ] Slightly Disagree
   - [ ] Mostly Disagree
   - [ ] Completely Disagree

29. Please, indicate other sources that you use to improve your grammar in Spanish (If any):

   [ ]

   281
30. I pay a lot of attention to content when I write my "Ejercicios De Escritura Libre" (EDEL).

- Always
- Very Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

31. I pay a lot of attention to grammar when I write my EDELs.

- Always
- Very Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

32. I try to write less complex sentences or ideas to avoid grammar errors.

- Always
- Very Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

33. I can write my Ejercicios de Escritura Libre (EDEL) faster than at the beginning of the semester (write more in less time).

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

34. How could the corrective feedback on your EDEL be improved?
35. What are you studying?

- [ ] Major in Spanish
- [ ] Minor in Spanish
- [ ] None of the above
APPENDIX D.2 CONTROL (NO-WCF) GROUP

1. Writing and revising my Ejercicios de Escritura Libre (EDEL) helped me notice my grammar errors.
   - Completely Agree
   - Mostly Agree
   - Slightly Agree
   - Slightly Disagree
   - Mostly Disagree
   - Completely Disagree

2. Writing and revising my Ejercicios de Escritura Libre (EDEL) makes more aware of my grammar errors in Spanish
   - Completely Agree
   - Mostly Agree
   - Slightly Agree
   - Slightly Disagree
   - Mostly Disagree
   - Completely Disagree

3. I know how to correct the grammar errors that I make in my EDELs.
   - Always
   - Very Frequently
   - Occasionally
   - Rarely
   - Very Rarely
   - Never

4. I am able to correct the errors that I make when I write and revise my EDELs.
   - Always
   - Very Frequently
   - Occasionally
   - Rarely
   - Very Rarely
   - Never
5. I ask the instructor when I revise my EDELs and I do not know how to correct a grammar error.

- Always
- Very Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

6. I make an effort not to make the same grammar errors when I write the EDEL.

- Always
- Very Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

7. I try to remember the errors I have made before when I am writing the EDEL.

- Always
- Very Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

8. I think it is important to have my errors corrected by the instructor.

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

9. I want to write more accurately in Spanish.
10. I want to improve my Spanish.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

11. I try to improve my grammar because I want to obtain a good grade in my coursework.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

12. I have learned from the errors I make when I revise my "Ejercicios de Escritura Libre" (EDELs).

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

13. I have learned new grammar concepts from the revisions I have made in my EDELs.

- [ ] Completely Agree
- [ ] Mostly Agree
14. The revisions have helped me obtain better command of grammar concepts that I already knew.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

15. Revising my EDELs has helped me learn grammar concepts that I previously did not understand.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

16. I believe that revising my EDELs has helped me decrease the frequency of some of the grammar errors I make.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

17. If so, indicate the error types that has been decreased as a result of the revisions you have made in your Ejercicios de Escritura Libre (EDEL).

- [ ] Regular Gender Agreement (i.e., la persona)
- [ ] Irregular Gender Agreement (i.e., el problema, el lugar)
- [ ] Number Agreement (i.e., todas las personas)
- Conjugations (i.e, ellos tienen)
- Omission of Articles (i.e, Las personas tienen…)
- Omission of 'que' (i.e., Pienso que no es necesario)
- Omission of "a" (i.e., ayuda a la gente)
- Subjunctive (i.e., no creo que sea un problema…)
- Spelling
- None of the above

18. Please indicate any other type of errors that you think you have reduced:

19. I feel anxiety when I have to revise my EDELs

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

20. I feel overwhelmed with the number of errors that I have to self-correct.

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

21. The time I spent revising my "Ejercicios De Escritura Libre" (EDEL) is worthwhile given how much I learned from it.

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree
22. I think the amount of grammar correction in the course is enough.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

23. I gain knowledge about grammar from reading texts in Spanish.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

24. I gain knowledge about grammar from a Spanish grammar book.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

25. I gain knowledge about grammar from other courses I am taking.

- [ ] Completely Agree
- [ ] Mostly Agree
- [ ] Slightly Agree
- [ ] Slightly Disagree
- [ ] Mostly Disagree
- [ ] Completely Disagree

26. I gain knowledge about grammar from interacting with other speakers of Spanish.
27. I gain knowledge about grammar from online activities.

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

28. I can write faster in Spanish than at the beginning of the semester.

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

29. Please, indicate other sources that you use to improve your grammar in Spanish (If any):


30. I pay a lot of attention to content when I write my "Ejercicios De Escritura Libre" (EDEL).

- Always
- Very Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

31. I pay lot of attention to grammar when I write my EDELs.
32. I try to write less complex sentences or ideas to avoid grammar errors.

- Always
- Very Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

33. I can write my Ejercicios de Escritura Libre (EDEL) faster than at the beginning of the semester (write more in less time).

- Completely Agree
- Mostly Agree
- Slightly Agree
- Slightly Disagree
- Mostly Disagree
- Completely Disagree

34. How could the revision process of your EDEL be improved?

35. What are you studying?

- Major in Spanish
- Minor in Spanish
- None of the above
APPENDIX E: INDIVIDUAL ACCURACY TRENDS

(WCF AND NO-WCF GROUPS)

Canonical Gender Assignment Individual Group Accuracy Linear Trends (WCF group)

Canonical Gender Assignment Individual Group Accuracy Linear Trends (No-WCF group)
Canonical Gender Agreement Individual Group Accuracy Linear Trends (WCF group)

Canonical Gender Agreement Individual Group Accuracy Linear Trends (No-WCF group)
Definite Articles in Obligatory Contexts Individual and Group Accuracy Linear Trends (WCF group)

Definite Articles in Obligatory Contexts Individual and Group Accuracy Linear Trends (No-WCF group)