“WHY ARE THESE UNDERLINED?” DEPTH OF PROCESSING AND TYPE OF WRITTEN CORRECTIVE FEEDBACK IN L2 SPANISH COMPOSITIONS

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

By

Allison Marie Caras, M.S.

Washington, D.C.
June 5, 2017
“WHY ARE THESE UNDERLINED?” DEPTH OF PROCESSING AND TYPE OF WRITTEN CORRECTIVE FEEDBACK IN L2 SPANISH COMPOSITIONS

Allison Marie Caras, M.S.

Thesis Advisor: Ronald P. Leow, Ph.D.

ABSTRACT

There is an ongoing debate as to whether written corrective feedback (WCF) is effective for the improvement of adult second language (L2) writers’ accuracy. Ever since Truscott (1996, 1999) began arguing against grammar correction in L2 writing courses, researchers have challenged his position (i.e. below). Although most of these studies show WCF to be effective, results are mixed about which type is most favorable. Researchers have investigated unfocused (the provision of feedback on any type of error) versus focused (the provision of feedback on a few linguistic targets), i.e. Ellis, Sheen, Murakami, and Takashima (2008), in addition to direct (via crossing out and the provision of the correction), i.e. Bitchener (2008), metalinguistic (via codes and the provision of a key to decipher their meaning), i.e. Ferris, Liu, Sinha, and Senna (2013), and indirect (via indicating the location of the error) WCF, i.e. Bitchener and Knoch (2010b). Furthermore, there is a dearth of empirical evidence that can provide how L2 writers process WCF given that studies typically employ data collection procedures that do not provide concurrent information on processing.

To this end, the current study attempted to address these gaps, namely, how Beginning L2 Spanish learners process WCF during the revision stage of a composition. Think aloud (TA) protocols were employed. Participants were randomly assigned to one of three experimental conditions (direct, metalinguistic, indirect) or the control condition. They revised their compositions twice: once with the respective WCF (Draft 2) and once with the original
composition (Draft 3). Accuracy was measured via scores for the production of *ser* versus *estar* and the preterit versus imperfect. These scores were submitted to repeated measures ANOVAs to locate potential differences within and between groups regarding type of WCF.

Results revealed that participants processed at high, medium, and low levels in the direct and metalinguistic WCF groups while the indirect WCF participants processed at low and medium levels. Type of WCF had no differential effect on accuracy scores over time.

*Key words:* depth of processing, written corrective feedback, compositions
ACKNOWLEDGEMENTS

I have oh-so many reasons to thank my family for seeing me through this process. Thank you, Marie, Bobby, and Chris Caras for showing interest in my studies, teaching, and activities over the course of this program. You have tolerated my ups and downs, from nerding out about a project to just needing quiet. Thank you for the trips, concerts, dinners, and movies. Over the past five years, you have accommodated this grad student lifestyle of mine from a place of utter support and over the past thirty years, you have encouraged me and given me the freedom to forge my path. In my life’s most pivotal decisions from college, to moving to California (and back), and diving into a Ph.D. program, you never thought twice when saying, “Go for it!” You are the reason I am where I am. This is for you.

Thank you, Logan and MacKenzie Caras, for the videos and laughs. Namely, my favorite question from a younger MacKenzie, “Aren’t you a little old to be in school?” Logan, you continue to inspire me with your video production skills. I can’t wait to see what’s in store for you both.

Ron Leow, my mentor, my academic grandfather, my friend. I have the pleasure of writing this note to you because of a few happy coincidences: from unknowingly living in adjacent neighborhoods to meeting through your academic son, who also happens to be my favorite professor at American University. Thank you, Ron, for challenging me to speak and write with true intention and strive to produce bulletproof work. You have given me the tools and pushed me to be the best student, teacher, and researcher I can possibly be.

Cristina Sanz, my mentor, my friend. I recall speaking with you during your Bilingualism & Cognition seminar about an upcoming final project. You told me, “Allison, you’re a first-year. This is going to hurt, but you will be fine.” I’ll never forget the compassion and encouragement I
felt from you in that moment. I have carried those words with me throughout the program like a mantra as I’ve tackled each project: “This is gonna hurt, but you’ll be fine.”

Luis Cerezo Ceballos, my mentor, my academic father, my friend. I remember standing in my parents’ kitchen after I got home from my first class with you at American. I felt electrified, rejuvenated, and validated for making the choice to leave my job across the country, move home, and further my education. I described to my parents how I felt during your class that night by imitating myself: elbows on the desk, fingers interlaced to cradle my chin, smile from ear to ear, and eyes as big as saucers. Fast-forward a few months to a conversation in your office about the next step. I’d decided I wanted to continue my education even further...and the rest is history. It has been an honor to continue working with you, to publish with you, to present at conferences with you, and to call you a friend. Thank you for the dinners, the laughs, and on a serious note, thank you for ushering me into the greatest accomplishment of my life so far.

To my professors at Gettysburg College: Dusty Smith, you are one of the principal reasons I wanted to become a professor. The way you related to us and conducted class as a roundtable conversation, valued our opinions, and pushed us to say more through our writing and mean it. You also introduced me to one of my best friends, Kim Dana Kupperman. Kim, you are an inspiration and I always look forward to our dinners. Len Mundy, you are the reason I majored in sociology and wanted to develop more of an edge. VoonChin Phua, thank you for taking me under your wing. Your continued advice is greatly appreciated. Jennifer Bloomquist, you are the reason I landed on linguistics. Thank you for allowing me to take a small set of your data to learn with, and in turn for introducing me to a great friend, Jessie Cox.

Thank you to my Georgetown professors, Hector Campos, Thomas Walsh, Alfonso Morales Front, Elena Herburger, Rusan Chen, John Norris, John Davis, Lourdes Ortega, Jennifer
Sclafani, and Darlene Howard. Thank you also to my wonderful Director of Graduate Studies, Michael Ferreira; my excellent Department Chair, Gwen Kirkpatrick; and to my stellar Department Administrative Assistant, Kristen Hall. Thank you to Peter Janssens for kindly reserving the language lab for me, for checking in on my progress throughout the program, and for recognizing my love for teaching and inviting me to the Language Teachers’ Circle sessions. Also, thank you so very much to Luke Plonsky for your time and brainstorming with me about statistics.

To my professors at American University, Luis, of course, Núria Vilanova, Carmen Helena Ruzza, Brenda Werth, and Ana Serra. Jack Child, I’m so grateful to have met you and although we didn’t get to work together, I always thought it would have been an absolute honor. Rest in peace.

To my Georgetown classmates, colleagues, and friends: Celia Chomón Zamora, mi media naranja. You are my go-to. Thank you for always having time for me and listening when I “need to think out loud.” From sitting on the floor with Ángela and Katie when we couldn’t find seats in the ICC to do our Vietnamese syntax work to holding hands when I hit the button to submit my Tarsky’s World semantics portfolio. From auditing Ron’s seminars after we were done with coursework to getting so excited about filling in our planners that we missed when we tried to high-five each other. Whoops. Secret’s out. Thank you for the coffees, the lunches, the dinners, the Jude videos, the puns. Ángela Donate Velasco, you have become one of my best friends. You are a force of nature and an absolute pleasure to be around. Thank you for the laughs, for lending an ear and brainstorming with me, and for your utterly contagious enthusiasm. It’s impossible not to smile when you’re around. Katie Vadella, thank you for your friendship, for being an
upstanding colleague, and for baking (and sharing) so many delicious goodies! You embody professionalism and I learned so much from you.

To my other friends and colleagues at Georgetown, Sandra Martín (¡Hoya Salamanca!), Janire Zalbidea, Tim McCormick, Jorge Méndez Seijas, Cristi Killingsworth, Young-a Son, Pablo Camus Oyarzún, Natalia Curto García Nieto, Lu Maimone, Nick Pandža, Ellen Johnson Serafini, Annie Calderón, Jessie Cox, Marisa Filguera Gómez, Ariel Zach, Colleen Moorman, Silvia Marijuan, Dan Riggs, Chrissy Bistline Bonilla, Gaby DeRobles, Lucia Donatelli, Will Travers, Meagan Driver, it has been a pleasure taking classes and working alongside you. Thank you, also, to the literature folks who have a special place in my heart for their friendship and collegiality, Oscar Amaya, Porter O’Neil, Estefania Tocado Orviz, Inés Corujo Martín, Alfredo Poggi, Ross Karlan, Felipe Toro Franco, Antonio Diaz Oliva, Jafte Robles, Verónica Zacipa, Gabriel Villarroel, Jose Cornelio (the best “cube-mate” out there), Bohumira Smidakova, Yoel Castillo Botello, Anthony Perry, Ashley Caja, Denise Kripper, María José Eleonora Navia, Jovana Žujević, Mónica Vallin, Carolina Rodriguez, and last but certainly not least, Ana María Ferreira (¡Hoya Salamanca!).

To my Georgetown students, you energized me and brightened my days. Thank you for the balance you brought to taking courses and writing. Thank you to my participants, as well, for attending each session with such great attitudes.

Thank you to Janet Sale (mi Pazcita), James York, Sam Sack, the Kennerknecht’s, Matt Yergey, Alan Fuerstman, Adam Rodriguez, Brandon Rodriguez, Ryan Stieler, Hank Wedaa, and countless other friends and family who have checked in with me over the years. Thank you, Danny Schwartz, for your friendship and all the music breaks.
To those who had no idea they were a huge help in the process: my favorite bands and artists. Whether it was a run at the Beacon or the Warner; a show or festival at Red Rocks or Notodden, thank you for playing your hearts out and leaving everything on the stage. Thank you, Tedeschi Trucks Band, Allman Brothers Band, Gov’t Mule, Marcus King Band, SwampCandy, and dozens more. You will never know the revitalization you brought me throughout this process. Rest in peace, Butch Trucks and Gregg Allman.

Finally, to my future students. This dissertation is in large part for you, too. I can’t wait to meet you and work with you.
DEDICATION

To my family

x
# Table of Contents

Chapter I: Introduction

Statement of the Problem

Chapter II: Review of the Literature

Investigations of Types of Written Corrective Feedback in Instructed Second Language Acquisition

- Unfocused direct versus unfocused indirect versus unfocused metalinguistic feedback
- Unfocused direct versus unfocused metalinguistic feedback
- Unfocused metalinguistic feedback
- Unfocused direct feedback versus unfocused indirect feedback
- Unfocused indirect versus unfocused metalinguistic feedback
- Unfocused indirect feedback
- Unfocused direct feedback

Focused direct versus focused indirect versus unfocused direct versus unfocused indirect feedback

Focused direct feedback

Focused direct versus unfocused direct feedback

Focused indirect feedback

Focused direct versus focused metalinguistic feedback

Focused direct versus focused indirect feedback studies

Depth of Processing in Instructed Second Language Acquisition Research

Empirical research on depth of processing

The Current State of Written Corrective Feedback Research

The Future of Written Corrective Feedback Research
LIST OF FIGURES

Figure 1: Summary of Pilot Data Collection Procedure ............................................. 66
Figure 2: Pilot Accuracy Score Formula ................................................................. 67
Figure 3: Pilot Depth of Processing Score Formula ............................................... 70
Figure 4: Example of a Participant’s Computer Screen during Time 2 ..................... 90
Figure 5: Summary of Data Collection Procedure ............................................... 91
Figure 6: Ser versus Estar Scores by Written Corrective Feedback Type ............... 116
Figure 7: Preterit versus Imperfect Scores by Written Corrective Feedback Type .... 119
LIST OF TABLES

Table 1: Summary of Unfocused Direct versus Unfocused Indirect versus Unfocused Metalinguistic Feedback Studies ................................................................. 6

Table 2: Summary of Selected Unfocused Direct versus Unfocused Metalinguistic WCF Studies ........................................................................................................ 11

Table 3: Summary of Unfocused Metalinguistic Written Corrective Feedback Studies.................................................................................................................. 14

Table 4: Summary of Studies Comparing Unfocused Direct versus Unfocused Indirect Written Corrective Feedback ........................................................................ 16

Table 5: Summary of Studies Comparing Unfocused Indirect versus Unfocused Metalinguistic Written Corrective Feedback ................................................. 18

Table 6: Summary of Unfocused Indirect Written Corrective Feedback Studies ........ 22

Table 7: Summary of Unfocused Direct Written Corrective Feedback Studies ....... 25

Table 8: Summary of Studies Comparing Focused Direct versus Focused Indirect versus Unfocused Direct versus Unfocused Indirect Written Corrective Feedback ..... 27

Table 9: Summary of Selected Focused Direct Written Corrective Feedback Studies ...................................................................................................................... 34

Table 10: Summary of Studies Comparing Focused Direct versus Unfocused Direct Written Corrective Feedback ................................................................. 40

Table 11: Summary of Focused Indirect Written Corrective Feedback Studies ........ 43

Table 12: Summary of Studies Comparing Focused Direct versus Focused Metalinguistic Written Corrective Feedback ......................................................... 45

Table 13: Summary of Studies Comparing Focused Direct versus Focused Indirect Written Corrective Feedback .......................................................................... 48

Table 14: Pilot Groups’ Components ...................................................................... 59

Table 15: Pilot Linguistic Targets .......................................................................... 61

Table 16: Pilot Coding Scheme for Depth of Processing of Feedback on Copulas ..... 68

Table 17: Pilot Coding Scheme for Depth of Processing of Feedback on Adjective
Gender Agreement and Past Tense Aspect

Table 18: Pilot Mean Scores for the Accurate Decision to Produce Ser

Table 19: Pilot Mean Scores for the Accurate Decision to Produce Estar

Table 20: Pilot Mean Scores for the Accurate Production of Adjective Gender Agreement

Table 21: Pilot Mean Scores for the Accurate Production of the Preterit

Table 22: Pilot Mean Scores for the Accurate Production of the Imperfect

Table 23: Pilot Depth of Processing Scores

Table 24: Groups’ Components

Table 25: Average Time on Task per Written Corrective Feedback Type at Time 2

Table 26: Coding Scheme for Depth of Processing of Ser versus Estar and the Preterit versus Imperfect

Table 27: Depth of Processing Scores

Table 28: Examples of Participants’ Think Aloud Protocols on Ser versus Estar per Depth of Processing Level

Table 29: Depth of Processing Results for Ser versus Estar per Written Corrective Feedback Type

Table 30: High Depth of Processing and Awareness at the Level of +/- Understanding for Ser versus Estar

Table 31: Examples of Participants’ Think Aloud Protocols on the Preterit versus Imperfect per Depth of Processing Level

Table 32: Depth of Processing Results for the Preterit versus Imperfect per Written Corrective Feedback Type

Table 33: High Depth of Processing and Awareness at the Level of +/- Understanding for the Preterit versus Imperfect

Table 34: Examples of Participants’ Comments or Actions per Written Corrective Feedback Type

Table 35: Commentary per Written Corrective Feedback Type
Table 36: Correlations between Depth of Processing Score and *Ser* versus *Estar* Score ................................................................................................................................. 111

Table 37: Correlations between Depth of Processing Score and the Preterit versus Imperfect Score................................................................................................................................. 112

Table 38: Descriptive Statistics for *Ser* versus *Estar* Scores per Written Corrective Feedback Type ................................................................................................................................. 113

Table 39: Summary of Between-Subject Analysis of Variance for *Ser* versus *Estar* .... 114

Table 40: Post hoc Scheffé for *Ser* versus *Estar* Scores on Draft 2 ......................... 114

Table 41: Post hoc Scheffé for *Ser* versus *Estar* Scores on Draft 3 ......................... 115

Table 42: Summary of Within-Subject Analysis of Variance for *Ser* versus *Estar* ....... 115

Table 43: Within-Subject Effect Sizes for *Ser* versus *Estar* ........................................ 117

Table 44: Descriptive Statistics for the Preterit versus Imperfect Scores per Written Corrective Feedback Type ................................................................................................................................. 118

Table 45: Summary of Between-Subject Analysis of Variance for the Preterit versus Imperfect ................................................................................................................................. 118

Table 46: Summary of Within-Subject Analysis of Variance for the Preterit versus Imperfect ................................................................................................................................. 119

Table 47: Within-Subject Effect Sizes for the Preterit versus Imperfect ....................... 120

Table 48: Average Time on Task, Number of Errors, and Think Aloud Protocols for *Ser* versus *Estar* per Written Corrective Feedback Type................................................................. 132

Table 49: T-test Results for Average Time on Task and *Ser* versus *Estar* Scores ...... 133

Table 50: Average Time on Task, Number of Errors, and Think Aloud Protocols for the Preterit versus Imperfect per Written Corrective Feedback Type ................................. 135

Table 51: T-test Results for Average Time on Task and Preterit versus Imperfect Scores ................................................................................................................................. 136
Chapter I: Introduction

Statement of the Problem

Since the 1970s, feedback on writing in the second/foreign language (L2) classroom has been of great interest to instructed second language acquisition (ISLA) researchers. In particular, there is an ongoing debate as to whether or not written corrective feedback (WCF) is effective for the improvement of adult L2 learners’ accuracy in compositions. Ever since Truscott (1996, 1999) began arguing against the provision of grammar correction in L2 writing courses, an increasing number of researchers have challenged his position. Polio (2012) took a critical stance on the entire debate by zooming out to the larger picture of SLA theories and whether or not WCF fits into them. She argued that “written feedback simply promotes better writing and is not intended to facilitate acquisition” (p. 376), but did not explain what exactly constitutes “better writing,” although she highlighted that within SLA theories, it could be argued that WCF expands explicit knowledge. However, “explicit knowledge” was not further explained either. In a parallel publication, Bitchener and Ferris (2012) stated, “[t]he question of whether [WCF] can play a role in the L2 acquisition process is the most important question to be answered. If the answer is negative, all other questions cease to have importance” (p. 50). Although the issue remains unresolved, potentially because of terminological conflation (i.e. what is meant by “acquisition” versus “development” versus “learning”), ISLA researchers are diligently scrutinizing WCF.

Traditionally, WCF has been realized in two ways: direct and indirect, (i.e. Hendrickson, 1978; Ellis, 2009). Bitchener and Knoch (2010b) explain that direct corrective feedback has been defined as feedback that provides:

Some form of explicit correction of linguistic form or structure above or near the linguistic error. It may consist of the crossing out of an unnecessary
word/phrase/morpheme, the insertion of a missing word/phrase/morpheme, and the provision of grammar rules and examples of correct usage. (p. 209)

More recently, Bitchener and Storch (2016) revised the operationalization of direct WCF by omitting the last detail from its previous description. In other words, in order to be considered direct WCF, explanations of grammar rules cannot be provided.

Indirect WCF has been described as:
That which indicates in some way an error has been made but it does not provide a correction. It is typically provided in one of two ways: (1) underlining or circling an error or (2) recording in the margin the number of errors in a given line. (Bitchener & Knoch, 2010b: 209)

Bitchener and Knoch further specify that editing symbols and codes have been included in the indirect WCF operationalization (i.e. Ferris & Roberts, 2001), but they do not consider them as such “because [they supply] additional metalinguistic information about the type of error from a linguistic perspective” (p. 209). Bitchener and Storch (2016) explicitly distinguish between indirect WCF and metalinguistic WCF and clarify the ever-broadening classification of the latter by offering specific details as to the provision of this type of WCF:

This is usually done by giving each error a number and at the bottom of the page of text or at the end of the full text providing the metalinguistic explanation and example(s) beside the relevant number assigned to the error category in the learner’s text. [It] can also be provided in the form of a clue and usually an error code is used for this purpose. (p. 17)

A trend in the past decade or so has been to examine a few specific linguistic items in L2 learners’ compositions rather than multiple grammatical, lexical, mechanical, and stylistic
categories. In the literature, the former type of WCF provision is considered “focused” while the latter is “unfocused”. In Ellis’ (2009) typology of WCF, he stated that “[t]eachers can elect to correct the students’ errors, in which case the CF is unfocused. Alternatively, they can select specific error types for correction” (p. 102). The definition of focused feedback was later expanded by Ferris et al. (2013) to refer to linguistic forms and structures “either predetermined by the researchers for the study or based on individual writers’ needs” (p. 309). Additionally, although the term “unfocused” has also been referred to as “comprehensive” in some studies (i.e. van Beuningen, de Jong, & Kuiken, 2008, 2012), for the purpose of maintaining consistent terminological dichotomies, the term for this type of WCF will be maintained as “unfocused”.

In the line of unfocused WCF research, investigators have used various agreed-upon guides for error correction (i.e. Chandler, 2003) to provide feedback on dozens of error types such as punctuation, missing words, spelling, verb forms, and word order to name a few. Some empirical studies that have scrutinized unfocused WCF, beyond populations of English learners, have looked at high school and college-aged learners of L2 Dutch (e.g., van Beuningen, de Jong, & Kuiken, 2008, 2012), German (e.g., Vyatkina, 2010), and Spanish (e.g., Frantzen, 1995; Kepner, 1991). On the other hand, investigators in focused WCF research have examined a few discrete linguistic items and grammatical structures at a time such as “a” and “the” in the English article system (e.g., Bitchener, 2008; Ellis, Sheen, Murakami, & Takashima, 2008; Sheen, 2007), past conditionals and relative clauses in English (e.g., Suh, 2010), and prepositions and the past simple tense in English (e.g., Bitchener, Young, & Cameron, 2005).

Although some of the following researchers did not explicitly classify the WCF type they provided to their participants as focused or unfocused, respectively, it will be categorized as such, based on their descriptions. To this end, WCF studies will be reviewed within the
following categories and their respective combinations (for studies that compare types of feedback): unfocused direct, unfocused indirect, unfocused metalinguistic, focused direct, focused indirect, and focused metalinguistic.

This dissertation is organized into five chapters. The current chapter introduces the topic at large and provides a statement of the problem. The second chapter expands on the literature review from above, delves deeper into the current state of the research, and concludes with the rationale for the study and research questions. The third chapter describes the design and methodology of the pilot study and its modifications for the current study and then describes the methodology of the current study. The fourth chapter presents the results. Finally, the fifth chapter offers a discussion, limitations, ideas for future research, and a conclusion for the current study.
Chapter II: Review of the Literature

Investigations of Types of Written Corrective Feedback in Instructed Second Language Acquisition

Unfocused direct versus unfocused indirect versus unfocused metalinguistic feedback. In an early WCF study, Robb, Ross, and Shortreed (1986) assigned 134 English as a foreign language (EFL) students enrolled in a composition course in a Japanese university to one of four experimental groups: unfocused direct feedback, unfocused indirect feedback in the form of highlighting or the total number of errors per line noted in the margin, or unfocused metalinguistic feedback. They wrote five narrative compositions throughout the academic year (34.5 classroom hours) and accuracy was determined by error-free T-units (i.e. Hunt, 1965). Although the researchers do not make it clear, it appears as though all writing assignments and revisions were completed outside of class. Results yielded no differences among the groups and all participants showed improvement in mean scores over time.

In the second of two studies, Chandler (2003) addressed whether direct and/or metalinguistic WCF had a differential effect on accuracy. Participants were 36 low proficiency ESL learners from a music conservatory enrolled in a writing course. They wrote 40-page autobiographies in five, eight-page chapters. Chandler argued that the increased length would consequently increase the treatments’ effects. She provided four types of feedback to all participants in different orders. The feedback types were unfocused direct correction, unfocused coded indirect feedback via underlining + metalinguistic codes, metalinguistic codes only, and unfocused indirect feedback via underlining only. The metalinguistic codes consisted of, for example, “ww” for wrong word, or “art” to indicate either an incorrect article was used or was missing altogether. Participants received one of the four feedback types after each chapter and
the analysis only included data from 20 participants from an original total of 36. The 16 participants who were eliminated did not turn in revisions after feedback on each chapter was provided. Results showed accuracy improved the most after unfocused direct and unfocused indirect, via underlining, corrections were provided by the teacher, without codes on revisions.

As for accuracy on a new text, Chandler compared participants’ final drafts with the next assignment and found that those who received unfocused direct or unfocused indirect corrections on the last assignment outperformed the others. Table 1 shows a summary of these studies.

Table 1

Summary of Unfocused Direct versus Unfocused Indirect versus Unfocused Metalinguistic Feedback Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robb et al. (1986)</td>
<td>134 college learners of English</td>
<td>Lexicon, syntax, style, measured by error-free T-units</td>
<td>(1) Unfocused direct (2) Unfocused metalinguistic (3) Unfocused indirect (highlighting) (4) Unfocused indirect (number of errors per line in margin)</td>
<td>5 compositions, 1 revision each</td>
<td>None</td>
<td>1 academic year (34.5 hours)</td>
<td>No difference among groups, Gradual improvement in mean scores among groups</td>
</tr>
<tr>
<td>Chandler (2003) Study 2</td>
<td>36 college learners of low English enrolled in writing course</td>
<td>Holistic, measured by mean number of errors per 100 words</td>
<td>(1) Unfocused direct (2) Unfocused indirect + underlining + metalinguistic codes (3) Unfocused metalinguistic codes (4) Unfocused indirect + underlining</td>
<td>5 8-page chapters of an autobiography, 1 chapter at a time, 1 revision per chapter before writing next chapter</td>
<td>None</td>
<td>10 weeks</td>
<td>All groups improved on accuracy Groups 1 and 4 outperformed Groups 2 and 3</td>
</tr>
</tbody>
</table>
Strengths and weaknesses of unfocused direct versus unfocused indirect versus unfocused metalinguistic feedback studies. The strengths of these two studies was their attempt to add a plethora of information to this line of research. However, they compared different types of WCF at the expense of including control groups. Neither study included one nor did it justify a reason, but given that both experiments took place in the classroom, a control group was not practical. Additionally, Robb et al. (1986) did not specify where the composition and revision processes took place, nor the amount of time on task. If the participants composed and revised at home, then the researchers could not have controlled for external factors or time on task. Chandler (2003) reported time on task, but relied on participant self-reports for these data. Both studies failed to control for attention paid to or processing of the WCF.

Unfocused direct versus unfocused metalinguistic feedback. Other early WCF studies such as Lalande (1982) and Semke (1984) reported conflicting results when comparing unfocused direct feedback groups with unfocused metalinguistic feedback groups. Lalande’s (1982) 60 participants were college learners of intermediate German who received one of the two above-mentioned feedback types on text revisions and new texts. The first essay, which was a plot summary of a story that had been studied in class, served as the pretest. Participants had 60 minutes to compose and when they received their respective WCF the next class day, they had 50 minutes to revise. Over the course of the 10-week quarter, participants composed a total of five essays. The first, reviewed above, and last served as the pretest and the posttest, respectively. The posttest followed the same format as the pretest. As for the essays composed in between, participants had 45 minutes to write and the same amount of time to revise (50 minutes). Results showed that the unfocused indirect feedback group outperformed the
unfocused direct feedback group in accuracy. Other studies (Ferris, 2006; Semke, 1984) have reported no differential effect of feedback type on accuracy.

In Semke’s (1984) study, 141 college students of third-quarter German at a U.S. institution completed a pretest, which consisted of a 10-minute free-write wherein they wrote as much as they could in German in the allotted time. The free-write pretest was aimed at diagnosing participants’ accuracy and fluency. Next, as part of the pretest, participants completed a 10-minute, 57-item multiple-choice cloze test, which probed their overall proficiency. The written tasks were short journal entries on which they received either comments on content, unfocused direct feedback, unfocused direct feedback plus comments on content, or unfocused indirect feedback via codes on their compositions. Every group wrote nine entries over the course of ten weeks except the metalinguistic group because these participants rewrote their entries every other week instead of composing new entries. The same tasks from the pretest served as the posttest. Results indicated no improvement in accuracy in any of the conditions.

In Ferris’ (2006) quasi-experimental study, she collected a total of 146 essays from 92 college-level ESL participants over the course of 15 weeks. She provided three instructors with an error categories and codes chart to use for the respective WCF groups: unfocused direct and unfocused metalinguistic. The experimental tasks were four, three-draft persuasive essays, based on readings from class. Participants revised each essay with the WCF. Ferris measured accuracy over time by comparing essay one with essay four. Results showed accuracy improvement in both groups and no significant difference between the groups.

Hartshorn, Evans, Merrill, Sudweeks, Strong-Krause, and Anderson (2010) and Evans, Hartshorn, and Strong-Krause (2011) took WCF in a novel direction by taking a dynamic instructional approach, which covers a wider scope than previous studies in that they
experienced with learners’ classroom contexts. Within the dynamic instructional approach is dynamic feedback, which is defined by Hartshorn et al. (2010) as “feedback that reflects what the individual learner needs most, as demonstrated by what the learner produces” (p. 87). The current author has classified dynamic feedback as unfocused since the researchers provided feedback on an extensive amount of categories. In Hartshorn et al. (2010), participants were 47 learners in two intact advanced-mid ESL writing classes and in Evans et al. (2011) they were 30 learners in two ESL classes, but their proficiency was described as having met the university’s requirements for matriculation. In both studies, the researchers assigned one class as “dynamic” and the other as “traditional”, wherein the dynamic group wrote compositions for 10 minutes per class period on a variety of topics and the traditional group wrote four multi-draft essays. The dynamic feedback group received unfocused metalinguistic feedback via symbols with a key to decipher them, they kept tally sheets of their errors, and the class activities were based on the most common errors made by the participants. They rewrote their compositions until they were error-free. The traditional group received unfocused direct feedback on rhetorical elements and a variety of linguistic items in each draft of their compositions. The researchers administered a pretest and posttest wherein all participants wrote for 30 minutes. Accuracy was measured by error-free T-units on all compositions. Results yielded positive results in accuracy for the dynamic group over time.

In his unpublished quasi-experimental dissertation, Mubarak (2013) investigated accuracy in the compositions of 46 college learners of low-level ESL studying media at a university in Bahrain. Participants were randomly assigned to the following groups: unfocused direct WCF, unfocused metalinguistic + indirect WCF via underlining, and no feedback. The no feedback control group received comments on their “performance” (p. 100), a term the
researcher did not explain but based the decision to provide general comments on Bitchener et al. (2005). During the pretest at the beginning of the course, participants had 60 minutes to write a minimum of 120 words in an autobiographical essay. The treatments occurred over the course of 12 weeks, wherein participants pre-drafted, drafted, and post-drafted. All three drafting stages of each composition took place in the classroom. During the pre-drafting stage, the class brainstormed, read, and/or discussed a topic from the course textbook. During the drafting stage, participants wrote all of their course compositions during class time and the researcher/teacher monitored their activity, occasionally sitting down with a student and discussing her/his composition. At the end of class, they turned in their compositions and they were returned with WCF if applicable the next class period. During this next class period, the post-drafting stage occurred wherein participants had a researcher/teacher-led discussion about their common errors. The researcher/teacher distributed their corrected (or not) drafts to them and instructed them to pay attention to the WCF. They had between 10 and 15 minutes to revise. The immediate posttest occurred after the 12-week writing treatment and consisted of another 60-minute session of a minimum of 120 words in a new piece of autobiographical writing. The delayed posttest was administered 10 months later and participants repeated the same procedure as the pretest and immediate posttest, with a slightly altered autobiographical prompting topic. A mixture of analytical methods to measure accuracy were employed, namely error-free T-units and Ellis and Barkhuizen’s (2005) obligatory occasion analysis, which targets the percentage of correct usage of a linguistic item, in this case English articles, prepositions, and verb tenses, when they are required in a specific grammatical context. Instead of providing feedback on these discrete linguistic items, Mubarak used them as a post-hoc way to measure accuracy. Results showed no difference among the groups. Table 2 gives a summary of the above studies.
### Summary of Selected Unfocused Direct versus Unfocused Metalinguistic Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lalande (1982)</td>
<td>60 college learners of intermediate German</td>
<td>15 grammatical and lexical categories</td>
<td>(1) Unfocused direct</td>
<td>5 compositions written in class (60 minutes for compositions 1 and 5, 45 minutes for 2, 3, and 4), each revised the next class day (50 minutes)</td>
<td>None</td>
<td>10 weeks</td>
<td>Group 2 outperformed Group 1</td>
</tr>
<tr>
<td>Semke (1984)</td>
<td>141 college learners of beginning German</td>
<td>Any errors</td>
<td>(1) Comments on content</td>
<td>9 journal entries on topics of participants’ choice, Group 4 revised every other week instead of writing</td>
<td>None</td>
<td>10 weeks</td>
<td>No group improved</td>
</tr>
<tr>
<td>Ferris (2006)</td>
<td>92 college learners of English enrolled in writing course</td>
<td>Verbs, noun endings, articles, word choice, sentence structure</td>
<td>(1) Unfocused direct</td>
<td>All 3 drafts of the first and fourth out-of-class essay assignments</td>
<td>None</td>
<td>15 weeks</td>
<td>Both groups improved on revisions</td>
</tr>
<tr>
<td>Hartshorn et al. (2010)</td>
<td>47 college learners of English enrolled in a writing course</td>
<td>Any errors</td>
<td>(1) Unfocused direct</td>
<td>Group 1 wrote 4 multi-draft essays. Group 2 free-wrote for 10 minutes in every class, revisions repeated until error-free. Both groups wrote essays with prompts for 30 minutes each.</td>
<td>None</td>
<td>15 weeks</td>
<td>Group 2 outperformed Group 1</td>
</tr>
<tr>
<td>Mubarak (2013)</td>
<td>46 college learners of low English, studying media</td>
<td>Any lexical, grammatical, mechanical errors</td>
<td>(1) Unfocused direct</td>
<td>12 compositions (during class, no time specified), 1 revision each (10-15 minutes)</td>
<td>None</td>
<td>12 weeks</td>
<td>No improvement No difference among groups</td>
</tr>
</tbody>
</table>
Strengths and weaknesses of unfocused direct versus unfocused metalinguistic studies.

While little WCF research has been conducted for an L2 beyond English, Lalande (1982) and Semke (1984) contributed findings for L2 German, but not without methodological concerns. For example, Lalande’s control group was not a true control group. Instead, the current researcher reclassified it as a direct WCF group because of the feedback these participants received. Semke did not technically include a control group either, which is one of the most important aspects of empirical research. However, the group that received comments on content exclusively could be considered an experimental control group. Ferris (2006), more clearly than her predecessors, did not include a control group. Also, since her study lasted an entire semester, she did not mention any attempt to control for external factors. Additionally, Semke’s experimental groups produced an uneven amount of written texts during the experiment. Since the metalinguistic group produced just over half the amount of writing the other groups did, and revised these pieces of writing every other week while the other groups produced new texts, these groups are arguably not comparable. Lastly, Lalande claimed that “the structured nature of the rewriting activity...compelled students to engage in guided-learning and problem-solving activities” (p. 143), but he did not include process measures to provide evidence for attention-paid, at least, so this statement is unwarranted.

As for the dynamic approach, while intriguing, the instructional approaches combined with the feedback conditions in Hartshorn et al. (2010) and Evans et al. (2011) were too different to truly compare especially due to the level of intensity and frequency with which the dynamic groups practiced writing. On the other hand, these studies contribute to the existing literature in that there are few studies that incorporate varying instructional approaches into their work (cf. Frantzen, 1995).
Mubarak’s instructions to the participants to “pay attention to the corrections” (p. 102) would have been justified by including an online process measure. Also, his decision to wait until 10 months after treatment to administer the delayed posttest seems quite extreme and unnecessarily long, especially given the low proficiency level of his participants. It would be useful for comparison’s sake across studies to control for amount of time between treatments and assessment tasks. Despite its quasi-experimental design, Mubarak did not account for external variables that might have played a role in the participants’ performance at the time of the delayed posttest. Neither Mubarak nor the other authors employed online process measures to control for attention paid.

None of the studies reviewed above included information regarding time on task. Although they provided allotted time, information about how much time the participants actually took to complete the tasks was not given.

**Unfocused metalinguistic feedback.** In a multiple-case study, Ferris et al. (2013) qualitatively investigated 10 U.S. college students of English as a second language (ESL) enrolled in a basic writing course. Over the course of one semester, the participants wrote four persuasive essays in class. The researchers provided the participants with focused indirect feedback on the 3-4 most common or most concerning errors, as determined by the researchers, on each of their essays and gave them an error they could consult while making revisions. Each revision session was followed by one-on-one interviews, which included time for specific questions about their texts and/or errors therein. Each student consulted an ongoing errors chart, reflective of her/his individual work, in each interview session. Descriptive results showed that the participants did not necessarily improve their written accuracy, but rather their error patterns
changed throughout the semester, indicating that focused indirect WCF paired with explicit grammatical discussions was ineffective. Table 3 displays a summary of this study.

Table 3

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/Feedback type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferris et al.</td>
<td>10 college learners of English enrolled in basic writing course</td>
<td>3-4 most common errors per participant</td>
<td>All participants received metalinguistic WCF + teacher conferences</td>
<td>Essays written in Weeks 1, 6, 11, and 16, (50 minutes each) + revisions (20 minutes each)</td>
<td>None</td>
<td>16 weeks</td>
<td>No improvement</td>
</tr>
</tbody>
</table>

**Strengths and weaknesses of unfocused metalinguistic studies.** Ferris et al. (2013), while one of the only longitudinal WCF studies to date and qualitatively informative, did not quantify the changes in accuracy over time and hence their claim that “explicit feedback such as the WCF provided for this study may fall short of meeting students’ needs if there are no opportunities for follow-up discussion and clarification” (p. 323), needs to be viewed with caution. Ferris et al. (2013) did not include an experimental group that received WCF exclusively, that is, either direct or indirect WCF without additional metalinguistic feedback. Their participants, therefore, received two types of feedback and were not compared to any other feedback or control groups. Ferris et al. did not control for online process measures and therefore cannot account for attention paid to the WCF.

**Unfocused direct feedback versus unfocused indirect feedback.** Frantzen (1995) not only provided unfocused direct WCF and unfocused indirect WCF, but also observed the effects of instructional environment on participants’ new texts. She observed 44 college learners of
intermediate Spanish in two groups: (1) a grammar-focused class, wherein the teacher provided unfocused direct feedback on texts, and (2) a content-focused class with no grammar supplementation, wherein the teacher provided unfocused indirect feedback via circling errors. Participants in both classes wrote four in-class compositions with the same amount of time in between them throughout the semester. They also wrote five out-of-class, 250-word compositions. They received WCF on every out-of-class composition and the second and third in-class compositions. The first and fourth in-class compositions served as the pretest and posttest and the participants did not receive WCF. They responded to the same prompt in each wherein they described their most memorable experience. Over the course of the semester, both groups improved their grammatical accuracy in the new text.

In Chandler’s (2003) Study One, she divided 31 music conservatory learners of high intermediate/advanced English as a second language (ESL) into two groups. All of the participants gradually wrote 25-page autobiographies in five, five-page chapters during the semester, turning in one chapter at a time. They also wrote a book review, which was not included in the experiment. She elected to provide unfocused feedback on a total of 22 English grammatical, lexical, and stylistic error types. Both groups received unfocused indirect feedback via the underlining of errors, but Group one was required to correct their errors and hand in a revision and then received unfocused direct corrections from the teacher of each chapter while Group two was not instructed to hand in corrections of their errors until they had turned in their entire autobiography. Once the teacher provided unfocused direct corrections on chapters one and two, Group two corrected their errors on chapters three, four, and five and then received unfocused direct corrections from the teacher on these chapters as well. At this point, both
groups turned in final drafts of their autobiographies. Results showed that Group one outperformed Group two. Table 4 below provides a summary of these studies.

Table 4

Summary of Studies Comparing Unfocused Direct versus Unfocused Indirect Written Corrective Feedback

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frantzen (1995)</td>
<td>44 college learners of intermediate Spanish</td>
<td>18 grammatical categories</td>
<td>(1) Grammar-supplementation class + unfocused direct (2) No grammar-supplementation class + unfocused indirect (circling)</td>
<td>In-class compositions (40 minutes)</td>
<td>None</td>
<td>15 weeks</td>
<td>Groups 1 and 2 improved</td>
</tr>
<tr>
<td>Chandler (2003) Study 1</td>
<td>31 music conservatory learners of high intermediate/advanced English</td>
<td>22 grammatical, lexical, stylistic errors</td>
<td>(1) Unfocused indirect (underlining) at time of revision + direct for each revision (2) Unfocused indirect (underlining) for each revision + direct in one session at end of semester on all drafts before submitting final draft</td>
<td>1 25-page autobiography, divided into 5, 5-page assignments</td>
<td>None</td>
<td>14 weeks (24 hours)</td>
<td>Group 1 outperformed Group 2</td>
</tr>
</tbody>
</table>

Strengths and weaknesses of unfocused direct versus unfocused indirect studies. The studies reviewed above, while they expand the body of research by providing novel participant backgrounds and contexts, do not come without some methodological issues. Chandler’s Study One control group was not a true control group since the participants received treatment (i.e. Bitchener, 2012a; Truscott, 2004). WCF studies must include control groups to regulate their
treatments’ potential effects and to properly report their results. While Frantzen reported time on task, Chandler did not, nor did she state where the participants composed their chapters. Both studies failed to control for attention paid to or processing of the WCF. They could have employed online data collection procedures to do so and subsequently explain their findings more robustly.

**Unfocused indirect versus unfocused metalinguistic feedback.** Ferris and Roberts (2001) is the only study to the current author’s knowledge that has investigated unfocused indirect versus unfocused metalinguistic WCF. They observed 72 college-level, low proficiency ESL learners’ self-edits on an opinion piece after they divided them into two experimental groups: one that received metalinguistic codes and one that received indirect via underlining, and a control group. The participants completed all experimental tasks in the classroom. The pretest consisted of three components: 1) a 5-item questionnaire about participants’ previous experiences with grammar in the classroom and their opinions about their teachers’ methods regarding grammar instruction, 2) a correction test wherein participants located errors in sentences and used the researchers’ symbol key to label them and, in the next section, used the symbol key to label previously underlined errors in essay excerpts, and 3) a written response to a prompt based on a short reading and participants had 10-20 minutes to complete parts one and two of the pretest components, and no more than 50 minutes to write. The researchers returned the essays to the participants with the respective feedback to the experimental groups and no feedback to the control group two weeks later and they had 20 minutes to make the necessary corrections. The experimental groups outperformed the control group and there were no differences between the experimental groups’ performances. Table 5 below provides a summary of this study.
Table 5

Summary of Studies Comparing Unfocused Indirect versus Unfocused Metalinguistic Written Corrective Feedback

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferris &amp; Roberts (2001)</td>
<td>72 college learners of low English enrolled in a writing course</td>
<td>Verbs, noun endings, articles, word choice, sentence structure</td>
<td>(1) Unfocused metalinguistic (2) Unfocused indirect (underlining) (3) Control</td>
<td>Opinion essay prompted by a reading (50 minutes)</td>
<td>None</td>
<td>2 weeks</td>
<td>Groups 1 and 2 outperformed Group 3 No difference between Groups 1 and 2</td>
</tr>
</tbody>
</table>

Strengths and weaknesses of unfocused direct versus unfocused indirect studies. A strength of Ferris and Roberts (2001) is that they contribute to the current body of research in a novel way. However, they stated at the outset that “the examination of student processing of error feedback under controlled conditions provides information that can subsequently be applied to the study of longer-term student progress” (p. 162), but they did not employ online procedures to capture their participants’ cognitive processes and therefore cannot make any claims about processing of WCF.

Unfocused indirect feedback. In Fathman and Whalley’s (1990) study, participants were 72 college learners in intermediate ESL writing courses. In class, they had 30 minutes to write a composition about an eight-picture sequential story followed by an oral summary. The researchers randomly assigned the participants to one of the following groups: no feedback; unfocused indirect grammar feedback only as indicated by underlining any grammatical errors; content feedback, i.e. “Interesting narration” (p. 182); and finally, a mixture of unfocused indirect grammar feedback and content feedback. After a few days, teachers returned the
compositions with the respective feedback (or not) to the participants and then they revised their texts for 30 minutes in class. Accuracy was calculated based on the number of grammatical errors in each composition. Results showed accuracy improved among participants who received unfocused indirect feedback or a combination of unfocused indirect feedback and content feedback.

Years later, Ashwell’s (2000) results corroborated Fathman and Whalley’s findings in that the unfocused indirect feedback groups outperformed the control group. In his study, he provided WCF in different orders per experimental group on three drafts to 50 participants, enrolled in two English-language writing classes at a university in Japan, divided into four groups: Group one received content feedback on the first draft and unfocused indirect feedback on the second draft; Group two received unfocused indirect feedback on the first draft and content feedback on the second draft; Group three received both content and unfocused indirect feedback on both drafts; and Group four served as a no feedback control group. Indirect feedback was realized via underlining errors and content feedback pertained to organization, paragraphing, and other general comments. The experimental essay was composed as part of the writing course and was the third essay of the term. Participants had composed essays in the same 3-draft process two times prior to the experiment. Ashwell calculated an accuracy rating for each draft of the experiment. The calculation was made by dividing the total number of errors into the total number of words in a given draft. A content score was calculated, based on the content comments given and a rubric devised by the researcher, by L1 English raters for each draft. A two-way ANOVA followed by a post-hoc Scheffe test reflected significant accuracy improvement in all the three experimental groups and not the control group, but there was no
difference between the experimental groups. No significant difference was found among groups in terms of content score over time.

Truscott and Hsu (2008) found similar results from the pretest to the immediate posttest in their study along this vein. However, since the improvement was not maintained by their experimental group from the immediate to the delayed posttest, they were still able to claim that WCF is not effective in aiding participants’ “learning as measured by performance on a new writing task” (p. 299). A group of 47 graduate students enrolled in a writing seminar at a university in Taiwan was split into an experimental group and a control group. The experimental group received unfocused indirect feedback and the control group received no feedback at all. In class during the twelfth week of the course, all participants had 30 minutes to write a narrative, Narrative 1, based on 8 sequential pictures representing a story of friends going to a birthday party. The next week, the experimental group received WCF in the form of red underlines and the control group received no feedback. They all had 30 minutes to revise their narratives, Narrative 1 revisions. One week later, all participants received a new sequential picture-description task of 8 pictures depicting the story of a boy who shoplifts and goes to jail. Once again, they were given 30 minutes to write their narratives, Narrative 2. The researchers calculated an error score for each narrative (Narrative 1, Narrative 1 revisions, and Narrative 2) by dividing the total number of errors into the total number of words. One researcher taught the writing course and the other did not, so the one who did not teach the course marked the narratives. She marked grammatical and spelling errors primarily but if she saw fragments or the misuse of punctuation she marked those errors as well. ANOVAs showed a significant difference between the experimental and control groups at the time of Narration 1 revisions, in which the participants who received unfocused indirect WCF lowered their error scores. However, at the
time of Narrative 2 there was no difference between the groups’ scores. Furthermore, the overall error scores in Narrative 2 were higher than the error scores in Narrative 1 which led the researchers to believe the two picture-description tasks were not comparable. They investigated the issue by standardizing the scores and re-analyzing the narratives via an ANOVA and the Wilcoxon rank-sum test and found the same result: there was no significant difference in error scores at the time of Narrative 2 between the groups. Truscott and Hsu suggest that even if learning took place among participants in the unfocused indirect feedback group, it was too “short-term to be detected by our Narrative 2” (p. 300). In conclusion, they argue that studies focusing on the revision process do not provide the opportunity for investigators to show evidence of learning and should include new texts in addition to revisions. Table 6 shows a summary of unfocused indirect feedback studies.
### Table 6

**Summary of Unfocused Indirect Written Corrective Feedback Studies**

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
</table>
| Fathman & Whalley   | 72 college learners of intermediate English | English verb forms, tenses, articles, agreement | (1) Unfocused indirect  
(2) Content comments  
(3) Unfocused indirect + content comments  
(4) Control | Picture description task (30 minutes), revisions (30 minutes) | None                        | A few days (not specified)                  | Groups 1 and 3 outperformed Groups 2 and 4 |
| Ashwell (2000)      | 50 college learners in English writing courses | English grammatical structures, lexical forms, and mechanical errors | (1) Content feedback on 1st draft; unfocused indirect on 2nd draft  
(2) Unfocused indirect on 1st draft; content feedback on 2nd draft  
(3) Content + unfocused indirect on 1st and 2nd drafts  
(4) Control | Essays (time not reported, completed at home); instructors’ time limit of 12 minutes per WCF session per participant | None                        | 1 academic year (28 sessions, 90 minutes each) | Groups 1-3 improved accuracy, no differences among them |
| Truscott & Hsu (2008)| 47 graduate students enrolled in a basic writing seminar | English grammatical and spelling errors | (1) Unfocused indirect  
(2) Control | Narrative story based on picture prompts (30 minutes), revisions (30 minutes) | None                        | 2 weeks                      | Group 1 outperformed Group 2 on the immediate posttest but not on the delayed posttest |
**Strengths and weaknesses of unfocused indirect studies.** One strength of these studies is that they all addressed accuracy in text revisions, providing the reader with some confidence in comparability; however, in order to improve the field’s understanding of the effectiveness of unfocused indirect WCF, researchers will need to test various feedback types across new texts. Fathman and Whalley (1990) arguably overstepped their bounds when they concluded that given their findings, teachers could start to “decide how, when, and to what extent they will respond to student errors” (p. 187). This suggestion is a bit ambitious and one needs to consider more than the findings of one study before making curricular decisions. Furthermore, none of these studies collected data on how the participants processed the WCF while they were exposed to it so for Fathman and Whalley to suggest that teachers decide how to provide WCF appears to be unwarranted. Until there is much more robust research on the matter, practitioners should not employ one method over another based on a paucity of robust empirical studies.

**Unfocused direct feedback.** In Kepner’s (1991) study, 60 U.S. college students of intermediate Spanish wrote eight journal entries, each consisting of a minimum of 200 words, at home as homework assignments every two weeks during one semester. The researcher and her two trained assistants provided either (1) unfocused direct feedback in the form of error corrections at the sentence level and an explanation of the rule or (2) content-related comments, all in green ink pen. Results showed that error corrections and metalinguistic rule explanations were ineffective for participants’ written accuracy improvement. Polio, Fleck, and Leder (1998) also found that the provision of unfocused direct feedback yielded no difference between the feedback group and control group.

Polio et al. (1998) measured accuracy in text revisions of 64 college learners of ESL over the course of a semester. They either participated in the control group or received unfocused
direct WCF, additional review, and editing instruction. During the second week of the semester, all participants wrote descriptive essays in response to one of two questions, their choice which one, and the researchers informed them that they would be graded. As a pretest, they had 30 minutes to complete the essay. In class two days later, the researchers returned the essays to the participants for revisions to be completed within 60 minutes. During the fifteenth week of the semester as a posttest, they chose a new prompt and wrote a descriptive essay in response for 30 minutes. Two days later they received their essays back for revisions to be completed within 60 minutes. Participants did not receive WCF on these essays. Instead, over the course of seven weeks the control group composed four journal entries per week. They did not receive WCF. On the other hand, each week the experimental group composed a journal entry, completed editing exercises on passages and grammar review of one focal grammatical component per week, wrote another journal entry, and revised one of the two entries from that week. Weekly focal grammatical points were subject-verb agreement, tense, passives and conditionals, modals, and word forms, with the exception of nouns and articles, which spanned two weeks instead of one. The researchers provided the experimental group with unfocused direct WCF on the journal entries as well as on the exercises. For analytical purposes, the researchers only recorded grammatical and punctuation errors and misuse of lexical items. Results showed that both groups improved in accuracy, in terms of error-free T-units, and there was no difference between the two groups. Table 7 provides a summary of these studies.
Table 7

*Summary of Unfocused Direct Written Corrective Feedback Studies*

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kepner (1991)</td>
<td>60 college learners of intermediate Spanish</td>
<td>All sentence-level errors</td>
<td>(1) Unfocused direct + explanation of rule (2) Comments on content</td>
<td>Journal entries (time not controlled, written at home)</td>
<td>None</td>
<td>12 weeks</td>
<td>No improvement</td>
</tr>
<tr>
<td>Polio et al. (1998)</td>
<td>65 college learners of English</td>
<td>English subject-verb agreement, verb tenses, passives and conditionals, modals, word forms</td>
<td>(1) Unfocused direct + editing instruction (2) Control</td>
<td>Essays (30 minutes) and revisions (60 minutes)</td>
<td>None</td>
<td>15 weeks</td>
<td>No improvement</td>
</tr>
</tbody>
</table>
Strengths and weaknesses of unfocused direct studies. The studies that showed WCF to be ineffective must be interpreted with caution, as they have some methodological issues. Kepner (1991), for example, did not include a control group or establish the participants’ existing written accuracy level via a pretest. Additionally, while she ensured a minimum word count in the participants’ journal entries, she did not enforce a maximum word count or even report the average length (i.e. Bitchener & Knoch, 2008). Lastly, and quite interestingly, she interpreted her findings to be ineffective, but the data show that error correction was indeed effective (i.e. Bitchener & Knoch, 2008; Ferris, 2004). As for Polio et al. (1998), the semester-long duration of the study raises the question of whether or not external factors played a role. However, in pedagogically-driven research, this is a risk researchers take. Overall, these studies contribute to the literature on less of an empirical level and more of a pedagogical level. It would still be of interest, though, to know how their participants processed the WCF through online process measures.

Focused direct versus focused indirect versus unfocused direct versus unfocused indirect feedback. Recently, Aghajanloo, Mobini, and Khosravi (2016) investigated how 120 intermediate ESL high school learners, ages 14-18, interacted with four types of WCF. The pretest and posttest were writing tasks, based on writing prompts from the Test of English as a Foreign Language (TOEFL). In addition to the pretest, participants’ proficiency was determined by the Nelson Proficiency Test. For the treatment, participants wrote five one-paragraph responses to particular topics relevant to their classes and the researchers provided them the respective WCF and scored their writing based on a rating scale of 1-30 points, including content, organization, vocabulary, cohesion, grammar, and form. Although time-on-task was not reported, the researchers mentioned that in the pilot study, it took participants 120 minutes to
complete the proficiency test and the pretest altogether. After 18 class sessions, they found unfocused direct WCF to be most effective.

Table 8 displays a summary of this study.

Table 8

Summary of Studies Comparing Focused Direct versus Focused Indirect versus Unfocused Direct versus Unfocused Indirect Written Corrective Feedback

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aghajanloo et al. (2016)</td>
<td>120 high school learners of intermediate English</td>
<td>Not reported</td>
<td>(1) Focused direct (2) Focused indirect (3) Unfocused direct (4) Unfocused indirect</td>
<td>Written responses to prompts</td>
<td>None</td>
<td>18 class sessions (No time period reported)</td>
<td>Groups 2, 3, and 4 outperformed Group 1</td>
</tr>
</tbody>
</table>

**Strengths and weaknesses of focused direct versus focused indirect versus unfocused direct versus unfocused indirect feedback studies.** While the study reviewed above contributes more research on high school learners, it does not come without its caveats. Namely, the researchers did not include a control group or specify to the reader, for replicability purposes, what linguistic targets were measured in the focused WCF groups. Additionally, they stated that “corrective feedback in L2 writing research is a controversial issue and an important factor in learning” (p. 29), but they did not include process measures to contribute to this statement. Lastly, the discrete breakdown of the timeline was not clearly specified, but the entire study lasted over 18 class sessions, which seems like quite a long time and external factors could have played a role in addition to the WCF.
**Focused direct feedback.** Another one of the most popular approaches to investigating accuracy in WCF research is through focused direct feedback. Bitchener et al. (2005) aimed their direct WCF at the three most common error types that emerged in the pretest task (English prepositions, the past simple tense, and definite articles) and therefore provided feedback only on these errors in four 250-word writing tasks over 12 weeks. They operationalized direct feedback as the provision of explicit corrections above underlined errors on 53 post-intermediate English for Speakers of Other Languages (ESOL) learners’ writing. Participants were divided into three groups from intact classes. Group one received focused direct WCF and had five-minute conferences with one of the researchers to ask clarification questions regarding the feedback they received. Group two received focused direct WCF only. Group three did not receive WCF but they did receive content comments, i.e. overall organization. Results showed no significant improvement in accuracy among the three linguistic items when they were grouped together. However, the experimental group that received a combination of direct feedback and a five-minute conference significantly outperformed the other groups on accuracy of the simple past tense and definite articles in new pieces of writing.

In a subsequent study, Bitchener (2008) investigated the effect of corrective feedback on accuracy in the written production of two functions of the English article system. Participants were 75 low intermediate ESL students in two private schools and they completed picture description tasks. Group one received direct feedback on each targeted category, which was the correct use of the referent “a” or “the”, and a written metalinguistic explanation and a 30-minute oral metalinguistic explanation. Group two received direct feedback and a written metalinguistic explanation, while group three only received direct feedback, and group four did not receive any type of feedback. The pretest was a picture description task wherein learners scored an average
of 57.73% accuracy on the use of “a” and “the”, showing they did not have full control of the use of these items. Two weeks later, they received feedback, or not, depending on their assigned condition and the immediate posttest was a second picture description task. One week later, this piece of writing was returned to the participants of the various experimental conditions. Two months later, the delayed posttest was realized in the form of a third picture description task. Learners did not receive feedback on this task. Accuracy was calculated as a percentage of correct use of “a” and “the” when they were required in the context. Results showed a significant improvement in the accurate production of English articles from the pretest to the posttest and in the maintenance of it from the posttest to the delayed posttest among all three experimental groups. Successive replications by Bitchener and Knoch (2008, 2009a) corroborated these results with a different participant pool of 144 learners from two private language schools and a university and 39 learners from a university, respectively. Soon after, Bitchener and Knoch (2009b, 2010a) extended their studies to include five writing tasks over a 10-month period. Participants were 52 low intermediate ESL learners in four intact classes at a university. Again, the experimental groups outperformed the control group consistently over time.

Sheen (2007) posited that she would find evidence for the acquisition of English articles through WCF in intermediate ESL learners. She assigned 111 participants in intact classes to one of three groups: focused direct WCF, direct metalinguistic written explanation, or the control group. During the two treatment sessions, the control group had class as usual and the participants in the experimental groups read a story silently, the teacher reviewed vocabulary items and the story outcomes with the class, and then they handed the teacher their copies of the story. The teacher then read the story aloud once while participants jotted down key words. Rationale for having the students read the text and then listening to it while taking notes was not
provided. Finally, participants rewrote the story and the researcher collected all the written retellings. Before the next session, she marked the retellings with the respective WCF on English article errors and a handful of non-article distractor errors. A few days later, participants were instructed to review their corrected texts silently for five minutes. In order to measure acquisition, Sheen administered a battery of tests: a speeded dictation test (8 minutes), a picture description writing test (12 minutes), an error correction test (15 minutes), and a language analytic ability test (20 minutes). They were administered as the immediate posttests and delayed posttests and an alternate version of the picture description task was used from the immediate posttest to the delayed posttest because participants in the pilot study complained about having to repeat the same writing task. Control group participants completed the tests but did not participate in the story retelling tasks. Results showed evidence of improvement, as measured by percentage of target-like utterances (TLU), in text revisions from the pretest to the immediate posttest, but not from the immediate to the delayed posttest.

Shintani, Ellis, and Suzuki (2014) scrutinized the indefinite article “a” or “an” and the English hypothetical conditional. While the English article system studies in Bitchener and colleagues’ repertoire involved participants from diverse linguistic backgrounds, the participants in Shintani et al. (2014) all spoke Japanese as a first language (L1). These participants were enrolled in a pre-intermediate English course at a university in Japan. They were randomly assigned to one of four experimental groups or the control group. Group one received direct WCF. Group two received a handout with metalinguistic explanations of the target structures. Group three received a metalinguistic explanation handout and the opportunity to revise their text and Group four received direct WCF and the opportunity to revise their text as well. Group five served as a control group and only composed new texts. On the first day of the experiment,
all participants took notes as they listened to a recording of a text two times and rewrote the story. The researchers elected to use a dictogloss task because it would prompt the participants to use the target linguistic structures in their written retellings. The researchers also created an interview group of 10 participants. Five participants received direct WCF and five participants received a metalinguistic explanation handout. They participated in the same experimental procedure as the other groups but the researchers interviewed them after the second dictogloss task to ask them about the feedback they had received on the first retelling and what they were thinking about while they wrote the second story retelling. All participants, except for those in the control group that did not receive any feedback, received feedback on the first piece of writing only. One week later (Time 2), participants that received direct WCF without the opportunity for revision had 5 minutes to review the corrections on their texts. The researchers collected the corrected texts and the participants completed the next dictogloss story retelling. Participants that received direct WCF and the opportunity for revision followed the same procedure as those who only received direct WCF but with the addition of a 20-minute revision period in which they were permitted to look at their corrected texts while they rewrote and then handed in their original texts and revised texts to the researchers before beginning the second task. As for the metalinguistic explanation groups, at Time 2 they were given the handouts to study and had 5 minutes to look over their original texts. The researchers collected the original texts and handouts and the metalinguistic explanation group without revision completed the second task. The revision group also handed in their original texts and handouts and then had 20 minutes to make corrections on their texts before moving on to the second dictogloss task. Finally, the control group received their original texts, had 5 minutes to look them over, turned them back in to the researchers without rewriting, and proceeded to the second task. One week
later (Time 3), all participants completed a third dictogloss task. The researchers calculated accuracy scores for each of the target structures. While they scored the texts, they eliminated 33 participants from analysis because they had not produced a single use of a hypothetical conditional sentence and therefore did not receive direct WCF on this structure or apply the metalinguistic explanation handout to their texts. In the end, data from 140 participants remained for analysis. All experimental groups significantly improved their accuracy of the English hypothetical conditional structure from the first text to the second text, and although their scores decreased by the time of the third text, the loss was not significant but they did not outperform the control group, which had non-significantly increased its accuracy scores over time. The only group to maintain its accuracy score and outperform the control group over the three weeks was the direct WCF + revision group. There were no differences between groups with regard to the English indefinite article. The researchers generalized the sample of interviews in order to offer an explanation by suggesting that “learners paid little attention to the indefinite article in their text reconstructions” (p. 124). However, retrospective interviews do not offer insight into what the participants are doing while interacting with the feedback so such a suggestion is unwarranted since they cannot claim how and when attention may have been paid to the indefinite article.

More recently, Stefanou and Révész (2015) added to the growing body of focused direct studies aimed at English referential articles by contributing research on L1 Greek, L2 English learners. They divided 89 high school learners into three groups: focused direct, focused direct + written explanation at the top of the page, and control. The participants read a story and completed a story retelling task, guided by pictures related to the content of what they read. The tasks were self-paced and participants spent an average of 20 minutes on each experimental task.
The next time they met, they reviewed their respective feedback for 5 minutes and then completed another story retelling task. Participants’ new texts, then, were used as measures of written production accuracy of the linguistic targets since they were never asked to revise their texts. The authors point out that revisions are not part of this particular classroom’s practices, so the participants were used to this type of approach to the writing process. The control group completed all the same tasks as the experimental groups, but they received spelling corrections instead of corrections on the target items. Results showed that the two experimental groups outperformed the control group on one of the article types and not the other, and there were no differences between the experimental groups.

Table 9 summarizes a selection of the above-mentioned studies.
Table 9

Summary of Selected Focused Direct Written Corrective Feedback Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitchener et al.</td>
<td>53 post-intermediate college learners of English</td>
<td>Prepositions, past participles, definite articles</td>
<td>(1) Direct + mini conference (2) Direct (3) Control</td>
<td>Essays (45 minutes)</td>
<td>None</td>
<td>12 weeks</td>
<td>Group 1 outperformed other groups on prepositions; Group 1 outperformed Group 2 on past simple tense; Group 1 outperformed Group 3 on definite articles; No significant difference among groups when all 3 targeted categories combined</td>
</tr>
<tr>
<td>Sheen (2007)</td>
<td>91 community college learners of intermediate English</td>
<td>Referential articles</td>
<td>(1) Direct (2) Direct metalinguistic explanation (3) Control</td>
<td>Story retelling tasks (5 minutes to review WCF, writing time not reported)</td>
<td>None</td>
<td>2 months</td>
<td>Groups 1 and 2 improved on revisions; Group 2 improved on new texts as well</td>
</tr>
<tr>
<td>Shintani et al. (2014)</td>
<td>140 college learners of pre-intermediate English</td>
<td>English hypothetical conditional and indefinite article</td>
<td>(1) Focused direct (2) Explanation on handout (3) Focused direct + revision (4) Explanation on handout + revision (5) Control</td>
<td>Dictogloss rewriting tasks (20 minutes)</td>
<td>None</td>
<td>1 month</td>
<td>Groups 1-4 outperformed Group 5 on accurate production of the hypothetical conditional from pretest to posttest; Only Group 3 maintained accurate production of the hypothetical conditional on delayed posttest</td>
</tr>
<tr>
<td>Stefanou &amp; Révész (2015)</td>
<td>89 high school learners of intermediate English</td>
<td>Referential articles</td>
<td>(1) Focused direct (2) Focused direct + explanation (3) Control</td>
<td>Story retelling task, guided with pictures (self-paced, average 20 minutes)</td>
<td>None</td>
<td>3 weeks</td>
<td>Groups 1 and 2 outperformed Group 3 on specific reference use but not generic No difference between Groups 1 and 2</td>
</tr>
</tbody>
</table>
**Strengths and weaknesses of focused direct studies.** Some of the noted limitations in the studies reviewed above include the inconsistency of total classroom exposure hours (i.e. Bitchener et al., 2005, wherein the experimental group that improved the most had received 20 hours of English instruction per week while the other groups had received 10 hours per week and 4 hours per week, respectively). The studies reviewed above provide insight into the possibilities of exploring the improvement of written production accuracy when employing a focused type of WCF. Although Sheen (2007) explicitly stated that she sought to find evidence for acquisition of the English article system, her approach was arguably flawed from the start in that she did not incorporate process measures in her design to allow her to capture participants’ online processes while interacting with the WCF. However, her work inspired extension studies (i.e. Ellis et al. 2008) that expanded the comparative focused and unfocused direct WCF literature. Stefanou and Révész (2015) point out some of their own limitations by acknowledging that the elicitation task used in their study was quite experimental and not as useful to the participants as a real-world task might have been and that because they elicited such specific uses of such specific linguistic targets, no generalizations can be extrapolated.

One strength of Bitchener et al.’s (2005) study is that, despite the range of total classroom instruction time each group received, the researchers controlled for total grammar instruction time. The researchers also made sure that the teachers did not review any of the target structures in class between data collection sessions. However, it could be argued that additional classroom hours could have allowed for more practice in general, thus giving Group one an advantage over the other groups. Another attribute of the study was that the researchers were not teachers in any of the participating classes and one researcher met with each of the participants in Group one for the individual conferences to maintain consistency in the content discussed.
When Bitchener (2008) and Bitchener and Knoch (2008, 2009a, 2009b, 2010a) refined their designs to investigate one target structure instead of three, they continued to operationalize their dependent variable as accuracy, not acquisition as observed in Sheen (2007). Accuracy rate was calculated by the percentage of correct usage of the target structure when required in the context.

A contribution to this strand’s methodology is Sheen’s (2007) use of multiple tests such as speeded dictation, a picture-description writing test, and an error correction test to triangulate participants’ knowledge of targeted structures. One concern, on the other hand, is that she might have placed additional cognitive demands on the participants by inadvertently testing their second language (L2) reading and listening comprehension abilities since their writing prompts were based on rewriting what they read and heard. In order to take a more critical look at these studies, Leow’s (1999) internal validity criteria checklist was employed.

Leow’s internal validity criteria are rooted in evaluating empirical studies regarding the role of attention. While the aforementioned studies do not claim to measure attention specifically, they do mention it and other cognitive processes while discussing accuracy or acquisition. For example, Bitchener et al. (2005) claim that “[t]he researcher drew particular attention to errors that were made in different linguistic environments” (p. 196). However, no process measures were employed in order to provide evidence that participants’ attention was indeed drawn to the target items. In broader “acquisition” terms, Bitchener (2008) suggests that future researchers “include several additional post-tests over a longer period of time…so that the ultimate value of written corrective feedback for acquisition can be determined” (p. 116). Additionally, Sheen’s (2007) first two research questions inquire as to whether WCF has an effect on the acquisition of English articles. However, she does not employ process measures to
probe participants’ internal processes such as attention while interacting with the feedback.

Shintani et al. (2014) describe the “acquisition” of their study’s targeted structures, but like their predecessors, no process measures were employed in order to collect online evidence of cognitive processes. They edged toward collecting process measures by distinguishing a small group of participants to be interviewed, but retrospective interviews are offline nonetheless and cannot capture online measures. Ellis et al. (2008) postulated that “a mass of corrections directed at a diverse set of linguistic phenomena…is hardly likely to foster the noticing and cognizing that may be needed for CF to work for acquisition” (p. 368). Leow’s (1999) review of attentional studies applies to these cases in that “researchers appeared to have relied totally on the performances of subjects on the post-exposure tasks to infer as to what they paid attention to while exposed to the L2 data” (p. 65). If researchers in the focused WCF strand wish to address learning processes, they must employ measures that allow them to do so.

Other criteria that were overlooked in all eight studies above were the reporting of time spent on experimental tasks, whether or not Hawthorne effects were likely, and whether the researcher ensured that all participants were on-task during the experiments. Sheen (2007) was the only researcher to report test reliability and Shintani et al. (2014) were the only researchers to report participant attrition. Although seven out of the eight studies collected data from intact classes, Bitchener and Knoch (2008) were the only researchers to mention that they randomized the assignment of each class to experimental and control groups. Lastly, Bitchener and Knoch (2009a) did not include a control group, which is the first criterion on Leow’s (1999) checklist. The exclusion of a control group is a design shortcoming and as Bitchener and Knoch (2008) themselves indicated, “…without a control group…there is no way of knowing whether the improvements in accuracy were a result of only the WCF treatment” (p. 411). The lack of a
control group excludes this particular study’s design from being considered empirically valid or robust.

**Focused direct versus unfocused direct feedback.** Ellis et al. (2008) and Sheen, Wright, and Moldawa (2009) explicitly used the labels “focused” and “unfocused” to classify and operationalize their feedback variables. Ellis et al. (2008), compared focused direct to unfocused direct WCF and found stronger and more promising results for focused direct WCF, this time across new texts. Participants were 49 intermediate ESL students who were majoring in industrial design. They were divided into three groups: focused direct feedback, unfocused direct feedback, and control. The pretest was administered immediately before the first writing task. It was an error correction test, borrowed from Sheen (2007), which consisted of 16 items with the correct and incorrect use of “a” (six items), “the” (six items), and four distractor items. Participants were instructed to identify and correct the errors. If they scored 90% or higher, they were removed from the data since Brown (1973) argued, “90% accuracy is commonly taken as the criterion level for acquisition of a grammatical feature” (p. 362). Participants completed four writing tasks over the course of 10 weeks. The tasks consisted of reading a short story accompanied by pictures (from Byrne, 1967), listening to their teacher’s (who was one of the researchers) explanation of key terms and re-reading of the story, the teacher then collected the stories, and participants re-wrote the story on a blank sheet of paper. No time limit was imposed. The experimental groups received focused or unfocused direct feedback and the control received no feedback; rather, general comments such as “Good!” or “Are they happy then?” (p. 359). Instead of revising their writing, they repeated the process with a new story, which was considered the immediate posttest. They repeated this cycle four times. Accuracy was scored based on Ellis and Barkhuizen’s (2005) obligatory occasion analysis. Results showed that all
three groups improved their accuracy from the pretest to the immediate posttest. The focused
direct feedback group also improved from the immediate to the delayed posttest, the unfocused
direct feedback group maintained its accuracy, and the control group’s accuracy declined.

Sheen et al. (2009) extended the previous two studies by adding a “writing practice
group”, which was considered the control group in the other studies and added a fourth group as
the (maturational) control, the participants of which only completed the assessment tests and
neither composed nor received any type of feedback. They found that all three experimental
groups outperformed the control group, but that the most effective treatment for accuracy was
focused direct WCF.

In a recent and quite large study, Shepherd, Daily O’Meara, and Snyder (2016) collected
writing samples from 279 college learners in L2 writing courses. The participating instructors
allowed their students, the participants, to choose from one of three feedback types, called
grammar agreements, that they would use for the duration of the semester. The grammar
agreements consisted of the following: 1) extensive unfocused direct WCF, 2) focused direct
WCF, and 3) minimal unfocused direct WCF. Since grammar was not graded in the courses,
these agreements were low-stakes and the participants chose based on their personal preference
without feeling pressured. The researchers found general improvements in grammar, mechanics,
and overall writing, but there were no significant findings. They did not offer the reader
explanations or examples of specific “grammatical errors” or “overall writing.”

Table 10 provides a summary of these studies.
### Summary of Studies Comparing Focused Direct versus Unfocused Direct Written Corrective Feedback

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
</table>
| Ellis et al. (2008) | 49 college learners of intermediate English | (1) English referential articles  
(2) English referential articles, past tense, prepositions, vocabulary | (1) Focused direct  
(2) Unfocused direct  
(3) Control | Story retelling tasks with guiding pictures (time not reported) | None | 10 weeks | Groups 1 and 2 improved from pretest to immediate posttest; Group 1 improved, Group 2 maintained, Group 3 declined from immediate to delayed posttest |
| Sheen et al. (2009) | 80 non-credit college learners of intermediate English | (1) English referential articles  
(2) English referential articles, copula ‘be’, regular past tense ‘-ed’, irregular past tense, temporal and locative prepositions | (1) Focused direct  
(2) Unfocused direct  
(3) Writing practice  
(4) Control | Story retelling tasks (time not reported) | None | 9 weeks | Groups 1-3 improved grammatical accuracy; Group 1 outperformed others in article accuracy |
| Shepherd et al. (2016) | 279 college learners enrolled in L2 writing courses at a U.S. university | Any grammar or 1-3 repeated errors, custom for each participant | (1) Unfocused direct, extensive  
(2) Focused direct  
(3) Unfocused direct, minimal | All writing assignments in the course (time not reported, completed at home) | None | 1 semester | No improvement |
**Strengths and weaknesses of focused direct versus unfocused direct studies.** As pointed out by the researchers, the sample size in Ellis et al. (2008) was indeed quite small. They also believe that their operationalization of focused and unfocused feedback might not have been successful. Another potential limitation of these studies, as offered by Ellis et al. (2008), is that although the researchers argue that it would be preferable for teachers to provide focused error correction on one error type at a time, they acknowledge that it is not feasible. They also caution that “a mass of corrections directed at a diverse set of linguistic phenomena…is hardly likely to foster the noticing and cognizing that may be needed for CF to work for acquisition” (p. 368). As Bitchener (2008) suggested, it is necessary to continue researching discrete linguistic items and adding to the literature in this way.

The author once again referred to Leow’s (1999) internal validity criteria checklist in order to evaluate the strengths of the methodological approaches in Ellis et al. (2008) and Sheen et al. (2009). Each study met the same criteria: Each had a control group, explicitly described the independent and dependent variables, exposed all groups to the same materials, allowed an equal total amount of time to all groups to complete the tasks, used alternate versions of the testing materials over data collection stages, reported the reliability of the dependent measures, and reported interrater reliability. One strength of Shepherd et al. (2016) is the contribution of observing learner agency in selecting WCF types.

Aside from the limitations the researchers themselves mentioned in their publications, there were emergent methodological concerns. For example, despite the fact that both Ellis et al. (2008) and Sheen et al. (2009) collected data from intact classes, neither set of researchers reported randomizing the classes to experimental or control conditions. Like their focused direct WCF cohorts, they did not report participant attrition, if Hawthorne effects were likely, or the
average time each group spent completing the tasks. They reported the allowed time, but not the actual time spent. For example, Ellis et al. (2008) reported that the experimental groups were instructed to “look over their errors and the corrections carefully for at least five minutes” but they did not supply further details (p. 359) or whether participants did indeed do so. As for Shepherd et al. (2016), the study may have been too ambitious in that L2 was not controlled. In other words, their participants came from a wide array of L2 writing courses and it is not plausible to make overarching generalizations with the results. Lastly, none of these studies employed online process measures to control for attention paid to the WCF.

**Focused indirect feedback.** Focused indirect WCF is another relatively unexplored approach that researchers have utilized to investigate written production accuracy. In an extension of Bitchener (2008), Bitchener and Knoch (2010b) used indirect feedback instead of direct feedback. Participants were 63 learners of advanced ESL in a U.S. university. They observed these participants’ written production accuracy of the English referential articles “the” and “a” once again, but established pretest scores of 83.08%-90.59% as the participants’ proficiency level was higher than that of the participants in Bitchener (2008). For the treatment, they completed picture description tasks for 30 minutes each, in class. Group one received indirect feedback via asterisks above their article errors and a written metalinguistic explanation with examples, group two received indirect feedback via circles around their article errors, group three received the same treatment as group one with the addition of an oral form-focused review, and finally, group four (control) received no feedback. Results showed that all three experimental groups outperformed the control group on the immediate posttest and groups one (indirect feedback via asterisks + written metalinguistic explanation) and three (indirect feedback via asterisks + written + oral metalinguistic explanations) outperformed groups two (indirect
feedback via circles) and four on the delayed posttest 10 weeks later. Although the participants started with a relatively high accuracy score, the researchers stated that groups that received indirect feedback and written and/or oral metalinguistic explanations were able to improve their accuracy even more over time. However, they did not run within-subjects analyses so their interpretation of these results must be taken with caution. Table 11 below provides a summary of this study.

Table 11

Summary of Focused Indirect Written Corrective Feedback Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitchener &amp; Knoch (2010b)</td>
<td>63 college learners of advanced English</td>
<td>English article system</td>
<td>(1) Indirect (asterisks) + written explanation with examples (2) Indirect (circles around errors) (3) Indirect (asterisks) + written explanation with examples + oral review (4) Control</td>
<td>Picture description tasks (30 minutes)</td>
<td>None</td>
<td>10 weeks</td>
<td>Groups 1-3 outperformed Group 4 on immediate posttest; Groups 1 and 3 outperformed Groups 2 and 4 on delayed posttest</td>
</tr>
</tbody>
</table>

**Strengths and weaknesses of focused indirect studies.** One notable, but controversial as mentioned above, contribution of Bitchener and Knoch (2010b) is that despite the fact that their participants had a relatively strong ability to produce English articles correctly in writing, they still improved and did not hit a “ceiling effect” (i.e. Frantzen, 1995). Seemingly keeping the missteps from their 2009a study in mind, Bitchener and Knoch (2010b) once again included a control group and reported that the same researcher administered the treatment for all
participants. They also extensively explained their independent and dependent variables, exposed the control group to the same materials as the experimental groups, provided different forms of the experimental tasks at the various phases of data collection, and reported interrater reliability. In all of their subsequent studies after 2009a, they ceased to report whether or not they controlled for exposure to the target structures outside the experiment. In fact, up to that point they had habitually reported that it was impossible to do so. In their 2009a study they provided justification for the absence of an external exposure check by stating that the “statistical measures were sufficiently robust to counter any effect that may have resulted from any such exposure” (p. 326).

The only study in which these researchers ever reported the random assignment of participants, in their case intact classes, to experimental or control conditions was in their 2008 study. They did not report participant attrition, whether or not Hawthorne effects were suspected, the amount of time participants spent on tasks, whether or not the researchers monitored them to make sure they were on task and following instructions, they did not report the reliability of their tasks, and they did not employ online process measures to control for attention.

**Focused direct versus focused metalinguistic feedback.** Like Bitchener et al. (2005), Bitchener, East, and Cartner (2010) focused their direct feedback on English singular and plural nouns and subject-verb agreement since they emerged as the top two most problematic structures in the pretest, which was a listening task followed by a written response. Twenty advanced ESL students at a tertiary institution in New Zealand wrote three blog entries for a maximum of 30 minutes after listening to prompts. One week later, the 10 participants in Group one received direct corrections in red type-face on their blog entries, on the target structures only. The other 10 participants who were in Group two received what Bitchener et al. (2010) called
metalinguistic feedback (i.e. Ferris & Roberts, 2001) since it took the form of abbreviated metalinguistic codes, i.e. “I do have a lots (n/s)…” (p. 6), where (n/s) stood for noun singular.

Group one was instructed to read over their errors and Group two was instructed to correct their errors next to each code. Once they finished, both groups listened to a new prompt and had 30 minutes to write another blog entry. One week later, they received feedback and did not have to write a new entry until four weeks later as a delayed posttest. Results reflected that both groups improved their accuracy in the production of singular and plural nouns, but only Group two (indirect feedback) improved in the production of subject-verb agreement. When the researchers combined the linguistic targets, results showed both groups improved and there was no difference between the two groups. Table 12 provides a summary of this study.

Table 12

Summary of Studies Comparing Focused Direct versus Focused Metalinguistic Written Corrective Feedback

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitchener et al. (2010)</td>
<td>20 tertiary institution learners of advanced English</td>
<td>English singular/plural nouns and subject-verb agreement</td>
<td>(1) Focused direct (2) Focused metalinguistic</td>
<td>Written responses to listening prompts (30 minutes each)</td>
<td>None</td>
<td>6 weeks</td>
<td>Both groups improved in singular/plural nouns; Group 2 improved in subject-verb agreement</td>
</tr>
</tbody>
</table>

Strengths and weaknesses of focused direct versus focused metalinguistic studies. In addition to its predecessors, this study shed more light on which type of WCF is effective in relation to linguistic target. Bitchener et al. (2010), while they did not include a control group, took WCF studies into the technological realm by providing WCF on blog entries. They did not
operationalize the use of technology as a variable, though, so no comparisons were made between computer and pencil-and-paper conditions.

Bitchener returned to his former design with his colleagues in the 2010 study by not including a control group. Additionally, as in previous studies, Bitchener et al. (2010) did not report participant attrition, the amount of time on tasks, or whether or not the researcher checked to make sure participants were following instructions during task completion. They also neglected to explicitly describe the dependent variable. It is also unknown as to whether or not the same researcher provided the treatments. Lastly, they did not employ online process measures to control for attention paid to or processing of the WCF.

**Focused direct versus focused indirect feedback studies.** In her 2010 dissertation, Suh randomly assigned 81 Korean college learners of EFL to one of five groups, either containing a think aloud (TA) protocol in order to observe their attentional processes or not, while interacting with feedback on English past counterfactual conditionals and object-of-the-preposition relative clauses. She pointed out that, while some previous WCF studies were based on learners’ attention, awareness, noticing, or other internal processes, they did not collect data in such a way that would show evidence for such occurrences. The treatment groups were focused direct feedback + TA, focused direct feedback – TA, focused indirect feedback + TA, focused indirect feedback – TA, and a control group (–WCF, –TA). The pretests were as follows: a story-retelling task to assess production (30 minutes), a multiple-choice test on the past conditional and a picture-drawing task for relative clauses to assess recognition of the target structures (15 minutes for both), and a second story-retelling task. All story-retelling tasks were based on readings in their L1 Korean and retold in their L2 English. As Suh pointed out, in order to ease the memory burden and control for vocabulary use as much as possible, she provided all participants guiding
sequential pictures and vocabulary after she took away the original story. A few days later, participants had 5 minutes to review the respective feedback on the second story-retelling task, and then completed a third story-retelling task. They repeated this step once more, for a total of four story-retelling tasks and three feedback sessions, and then completed the immediate posttests (production and recognition). The delayed posttests (again, production and recognition) were one week later and participants also took an exit questionnaire. Results showed the participants who received focused direct WCF outperformed the other groups on the English past conditional phrases. She also found that once she coded the TAs for levels of awareness, “[t]he majority of the participants who received direct feedback reported a high level of awareness while most participants receiving indirect feedback reported a low level of awareness” (p. 116).

Karimi (2016) observed 72 volunteers, ages 15-29, from a language institute for six sessions over the course of nine weeks. They administered the Oxford Placement Test to control for proficiency and a picture description task as the pretest. They were divided into focused direct, focused indirect, focused direct for the first three sessions and then focused indirect for the last three sessions, and control. They completed picture description tasks accompanied by key words in the margin. They repeated this over the six sessions, during class time, and the instructors provided the respective WCF on each task after each session. A one-way ANOVA showed that the experimental groups improved on the accurate production of the linguistic targets: English articles and the simple past tense. Furthermore, a post-hoc Scheffe test reflected that the direct group outperformed the indirect group and the direct + indirect group outperformed the indirect group. The researchers did not report what the control group did nor did they include the control group in their analyses. Table 13 summarizes these studies.
### Table 13

**Summary Studies Comparing Focused Direct versus Focused Indirect Written Corrective Feedback**

<table>
<thead>
<tr>
<th>Study</th>
<th>Participants</th>
<th>Target structures</th>
<th>Groups/WCF type</th>
<th>Task/time</th>
<th>Process measures</th>
<th>Overall timespan</th>
<th>Accuracy results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suh (2010)</td>
<td>81 college learners of low English</td>
<td>Past counterfactual conditional clauses and object-of-the-preposition relative clauses</td>
<td>(1) Focused direct + TAs (2) Focused direct – TAs (3) Focused indirect + TAs (4) Focused indirect – TAs (5) Control</td>
<td>Written story-retelling tasks (30 minutes), WCF review (5 minutes) before writing another story-retelling</td>
<td>TA protocols</td>
<td>4 weeks</td>
<td>Groups 1 and 2 outperformed Groups 3, 4, and 5 on past counterfactual conditionals</td>
</tr>
<tr>
<td>Karimi (2016)</td>
<td>72 learners of English at a language institute</td>
<td>Articles, simple past tense</td>
<td>(1) Focused direct (2) Focused indirect (3) Focused direct then focused indirect (4) Control</td>
<td>Picture description tasks (time on task not reported)</td>
<td>None</td>
<td>9 weeks</td>
<td>Group 1 outperformed Group 3 Group 3 outperformed Group 2</td>
</tr>
</tbody>
</table>

**Strengths and weaknesses of focused direct versus focused indirect studies.** Suh’s dissertation is quite methodologically sound, as she operationalized her study within more realistic terms than previous studies that claimed to measure acquisition of linguistic items, (i.e. Sheen, 2007) by employing online learner process measures, therefore becoming the only study to the current author’s knowledge that meets this criterion on Leow’s (1999) checklist. Furthermore, her dissertation met the most criteria on this checklist among all of the reviewed studies. She included non-TA groups to control for reactivity, which is defined by Leow (2002) as “the act of thinking aloud potentially affecting change in learners’ cognitive processes while
interacting with L2 data” (p. 117), which she did not find. Suh’s operationalizations of terms are arguably more realistic than previous studies that claimed to measure acquisition of linguistic items, (i.e. Sheen, 2007) in that she collected data on the construct of awareness via think aloud (TA) protocols and did not conflate the terminology. Furthermore, she included non-TA groups to control for reactivity (the interference of thinking aloud with the task at hand), which she did not find. She included a control group, explicitly detailed the independent and dependent variables, used the same materials with all participants, allowed equal time and reported all time spent on tasks among participants, administered the treatment to all participants, controlled for external exposure to target structures via a questionnaire, used alternate versions of tasks across the phases of data collection, reported interrater reliability, and as cited above, collected process measures. Finally, Suh (2010) randomly assigned her participants to experimental and control conditions, while Bitchener et al. (2010) reported the random assignment of participants to either of two experimental conditions.

Suh (2010) points out her own limitation of uneven participant attrition. However, the pretest scores across experimental and control groups were not statistically different and she maintains that “more even distribution of cells could strengthen the results” (p.150). She also mentioned that in future research she would have tested each linguistic target as a separate variable and how WCF might affect participants’ written production accuracy of each. As for Karimi (2016), there was a lot of variation within the participant pool in terms of age and language proficiency was not specified. He only pointed out that he used the Oxford Placement Test to control for homogeneity of participants’ proficiency. Also, he, the researcher did not provide WCF to the participants, but rather the instructors did. Although he reported that the experiment took place during class time, which was held once a week for 90 minutes, Karimi did
not report time on task per session. Since there were six treatment sessions over the course of nine weeks, external factors may have played a role in the results, apart from WCF.

Based on Leow’s (1999) internal validity criteria, Suh’s (2010) dissertation is the most robust of all the studies reviewed. In sum, she met 13 of the 17 stipulations. In keeping with Suh’s use of process measures to justifiably capture learners’ awareness, the author will take the research one step further by investigating learners’ depth of processing (DOP) while interacting with WCF. Hyland (2016) argued that TAs in writing research offer only a limited glimpse into learners’ thought processes because “many cognitive processes are routine and internalised operations, and therefore not available to verbal description” (p. 81). While many cognitive processes arguably are automatized indeed, in this supposition Hyland did not acknowledge that some writing studies investigate within the overarching writing process itself. They zero in on the processes involved in learning discrete linguistic targets, as seen in the WCF literature.

**Depth of Processing in Instructed Second Language Acquisition Research**

It is widely agreed upon that attention is an integral part of the learning process (e.g., Schmidt, 1990; Tomlin & Villa, 1994; Robinson, 1995). However, simply paying attention to the input does not necessarily mean the learner will take it in. The culprit is not a learner’s limited attentional capacity but a disruption in L2 processing (i.e., Leow 2015). Empirically, “protocols have revealed that awareness at the level of understanding is typically associated with hypothesis testing, rule formation, and conscious activation of prior knowledge...which are all indicative of a high level of processing” (p. 218). In other words, it is possible to process at a high level but not understand the input (awareness at the level of understanding).

In order to examine learners’ processing and processes, researchers have utilized concurrent data collection procedures such as TA protocols (e.g., Leow, 1997, 1998). One aspect
of L2 learners’ processing is how much cognitive effort is employed during processing, referred to as DOP, and the empirical studies that have examined DOP have done so at the initial stages of L2 learning and have pertained almost exclusively to learners’ processing of written input, (i.e. Leow, 2015). Empirical research shows that deeper, rather than shallow levels of processing, are beneficial to L2 learners’ performance on reading comprehension tasks, (e.g., Morgan-Short, Heil, Botero-Moriarty, & Ebert, 2012) and written tasks (e.g., Qi & Lapkin, 2001). For the purposes of empirical scrutiny, DOP must be captured online, while the learner is engaged in the experimental task.

**Empirical research on depth of processing.** ISLA researchers have employed various experimental conditions to investigate DOP, such as attentional conditions (Shook, 1994; Gass, Svetics, & Lemlin, 2003), glossing conditions (Rott, 2005), and form and meaning conditions (Morgan-Short et al., 2012), to name a few.

In their study on L2 learners’ written production, Qi and Lapkin (2001) used a picture prompt writing task to observe DOP. Participants were two adult learners of L2 English. Data were collected in three phases: First, participants used the L2 to write a response to a picture. Second, they compared their written versions with a reformulated version and were interviewed in order for the researchers to identify what they had noticed. Lastly, participants revised their written responses. The findings showed improvement in written production when the aforementioned stages were involved, indicating that the deeper the participants processed their compositions, the more improvement they showed over the course of the three stages of writing. One limitation the researchers pointed out was that they had two participants, so an expansion of this study with more participants would be necessary in order to extrapolate their findings.
In her study on word processing and glossing, Rott (2005) observed 10 L1 English, L2 German university-level students. They were randomly assigned to a multiple-choice gloss or a single-translation gloss condition and all of them read a text for meaning and thought aloud. Results reflected that learners who made stronger form-meaning connections outperformed learners who made weaker form-meaning connections, as evidenced by the TA, on a Vocabulary Knowledge Scale test and a L1 oral recall of the story in the text. It can be argued, then, that stronger and weaker form-meaning connections can be viewed in light of DOP, as seen in the different levels of processing in the learners, and that making stronger connections lead to superior performance on the posttests.

Morgan-Short et al. (2012) investigated 139 L2 learners’ reading comprehension by conceptually replicating Leow, Hsieh, and Moreno (2008), with the addition of a non-TA group to observe reactivity. Participants were randomly assigned into processing conditions, which were the same as in Leow et al., (2008), and then read a Spanish text while thinking aloud. The processing conditions were the following: read for meaning only; read for meaning and circle all occurrences of the lexical item sol (meaning sun); read for meaning and circle all occurrences of lexical item la (feminine singular definite article meaning the); read for meaning and circle all occurrences of lexical-grammatical hybrid item lo (clitic meaning it); and read for meaning and circle all occurrences of the grammatical item –n (third person plural verbal ending). Results showed that processing condition did not have an effect on comprehension and reactivity was not an issue. Of interest in the current study is that they found a positive correlation between deeper levels of processing and higher comprehension scores.

Calderón (2013) investigated DOP and aural input. She observed 24 L1 English learners of Spanish with varying levels of L2 proficiency. Participants listened to a passage in the L2
followed by an immediate posttest, which was a multiple-choice sentence completion test. Calderón used TA to measure depth of processing, among other variables such as levels of awareness while the participants took the experimental test. Results showed that higher L2 proficiency participants processed at lower levels overall than did lower L2 proficiency participants, indicating that the information had been automatized in the higher L2 proficiency learners, allowing them to access prior knowledge and therefore process at lower levels.

The Current State of Written Corrective Feedback Research

Researchers have approached WCF through a plethora of ways. They have investigated direct and indirect WCF either separately (i.e. Ashwell, 2000; Bitchener et al., 2010; Bitchener & Knoch, 2008, 2009a, 2009b, 2010a; Fathman & Whalley, 1990; Kepner, 1991; Truscott & Hsu, 2008) or by comparing them (i.e. Bitchener & Knoch, 2010b; Robb et al., 1986). They have explicitly explored focused and/or unfocused direct (i.e. Sheen et al., 2009; Shintani et al., 2013) as well as unfocused direct WCF (i.e. Evans et al., 2011; Hartshorn et al., 2010) and they have approached the provision of WCF from the larger picture of a broad range of elements (i.e. Chandler, 2003) to the smaller picture of WCF on specifically targeted linguistic items such as the English referential articles “a” and “the” (i.e. Bitchener, 2008; Ellis et al., 2008; Sheen, 2007).

The reader must be cautious when interpreting these results as Hyland and Hyland (2006) warned, “it is difficult to draw clear conclusions and generalizations from the literature as a result of varied populations, treatments and research designs” (p. 84, also see Ellis et al., 2008, p. 354.). Guénette’s (2007) review confronted this idea in that she examined the research design and methodology of WCF studies and reiterated that they are not essentially comparable. In a recent meta-analysis of WCF studies, Kang and Han (2015) point out that the findings of these
studies are still inconclusive and that the major issues are “(a) when and how written corrective feedback works and (b) what type of feedback strategy is effective” (p. 1).

The Future of Written Corrective Feedback Research

Overall, these studies have emphasized the importance of testing various types of WCF and its effects on different linguistic items in future research. They recognize that it is not practical for instructors to provide WCF on one linguistic item at a time, but it is important to research WCF in this way in order to decipher which type of feedback has the most beneficial outcomes in terms of written production accuracy of discrete linguistic items. In other words, it is not realistic in a classroom setting to spend an extensive amount of class sessions on a specific target structure, which is precisely what some researchers suggest by emphasizing a focus on discrete linguistic items across new texts (i.e. Bitchener, 2008; Ferris, 2004; Truscott, 1996, 1999). The value of measuring accuracy across new compositions is warranted, of course, but when one looks to the L2 classroom, Ellis et al. (2008) make a rather practical point that, “[t]eachers…may feel that they do not have the luxury of focusing exclusively on a single error when they correct their students’ written work. Clearly, if [W]CF is effective when it addresses a number of different errors, it would be advantageous to adopt this approach” (p. 367). It is unlikely that teachers spend more than a few class sessions on a given linguistic item based on their curricula. In sum, the research is moving toward measuring accuracy across new texts in order to make L2 acquisition or learning claims, but researchers might be getting ahead of themselves and leaving unanswered questions regarding text revisions behind.

One step toward filling in gaps in the research is to capture learners’ processes while they interact with the feedback. Bitchener and Knoch’s (2009a) suggestion to utilize retrospective interviews as a means of uncovering “the extent to which feedback facilitates the acquisition of
partially learned linguistic features through consciousness-raising as opposed to the acquisition of new linguistic knowledge” (p. 328) is a start, but retrospective interviews do not provide evidence of what the learners are (or are not) doing with the feedback during text revision or new text writing. For example, Simard, Guénette, and Bergeron (2015) employed retrospective interviews and a questionnaire to support their claim that “in order for WCF to lead to second language improvement, learners must first be able to not only correctly interpret the WCF but also understand the linguistic information provided through this feedback” (p. 233). However, their claims that participants formed incorrect hypotheses about the WCF can only be interpreted through what the participants said after-the-fact and not while interacting with the feedback. Bitchener’s (2012b) point that although constructs such as attention, awareness, and noticing have been tested, (without providing information for clarification either of example studies to which he refers or of the types of online data collection procedures employed), in the oral corrective feedback realm, these constructs have yet to be seriously addressed by WCF researchers. To the author’s knowledge, the only WCF studies that investigated learners’ online processes are Sachs and Polio (2007), although they only collected think aloud protocols in the reformulation condition and not the error correction condition, and Suh (2010).

Future research must attempt to capture learners’ cognitive processes as they interact with WCF. Additionally, the research must address more linguistic items in a variety of L2s other than ESL. To this end, the current author proposes to continue along the same concurrent path since online data collection procedures are most useful to observe learners’ cognitive processes at the point of encoding and decoding the feedback provided. Indeed, Pawlak (2016) points out that procedures such as TAs can assist researchers to “more fully understand the value of different corrective moves and their contribution at different stages of cognitive input
processing” (p. 721). In accordance with Leow’s (2015) model of the L2 learning process in ISLA, the author will investigate learners’ DOP while they interact with WCF. She will expand the current body of research by investigating Spanish L2 learners’ DOP, as evidenced by coded TAs, and by providing various types of WCF on different linguistic items in university-level Beginning Spanish classes.

**Rationale for the Study**

Based on the review of empirical studies, the current author attempted to contribute to the WCF literature by investigating how learners process during the revision phase of composition as well as different types of WCF and their potential effects on the accurate production of discrete linguistic targets in L2 Spanish learners’ compositions. By applying DOP to this strand of research, the current study furthers the discussion of the importance of online process measures when investigating WCF and learners’ written production accuracy. It must be noted that by no means is the current author claiming to measure “acquisition,” “development,” or “learning,” by analyzing these online processes, but rather use the evidence of these processes to support the claims regarding potential changes in accuracy. In other words, the study describes how participants process WCF and particular targets.

As Ellis et al. (2008) suggested, it is not practical in the classroom to provide feedback on discrete linguistic items one at a time in compositions. In order to judiciously scrutinize how learners process various types of WCF, the current author elected to follow the middle road and provide unfocused feedback, but to measure written production accuracy as it pertains to a selection of specific targets. The current study seeks to expand the ISLA literature base by informing how L1 speakers of English process WCF in the L2 Spanish writing process, if there is a relationship between levels of depth of processing and the accurate production of certain
targets, and whether particular types of WCF have an effect on the accurate production of these targets.

**Research Questions (RQs)**

The following research questions will be addressed:

RQ1: How do adult L2 Beginning Spanish learners process unfocused written corrective feedback on:

- RQ1a) *ser* versus *estar*?
- RQ1b) the preterit versus imperfect past tense aspects?

RQ2: Is there a relationship between level of depth of processing of written corrective feedback and adult L2 Beginning Spanish learners’ written production accuracy of:

- RQ2a) *ser* versus *estar*?
- RQ2b) the preterit versus imperfect past tense aspects?

RQ3: Does type of unfocused written corrective feedback (direct, indirect, metalinguistic) have a differential effect on adult L2 Beginning Spanish learners’ subsequent written production accuracy of:

- RQ3a) *ser* versus *estar*?
- RQ3b) the preterit versus imperfect past tense aspects?

If so, does the effect on accuracy on each respective target dichotomy last over 2 weeks?
Chapter III: Research Design and Methodology

This chapter addresses the pilot study design and methodology, linguistic targets, materials, procedure, scoring, coding, analysis, results, discussion, methodological issues, and modifications for the current study.

Pilot Study

The pilot study was administered in Weeks 14 and 15 of the Spring 2016 semester. The following research questions were addressed:

RQ1: Does type of unfocused written corrective feedback (direct or indirect) have a differential effect on adult L2 Beginning Spanish learners’ written production accuracy of:

  RQ1a) the copula ser?
  RQ1b) the copula estar?
  RQ1c) gender agreement in adjectives?
  RQ1d) the preterit past tense aspect?
  RQ1e) the imperfect past tense aspect?
  RQ1f) If so, does the effect on accuracy on each respective linguistic target last over 5 days?

RQ2: How do adult L2 Beginning Spanish learners process unfocused written corrective feedback?

Pilot design and methodology. Participants in the pilot study were 18 students of Non-Intensive Beginning Spanish (SPAN 001), in two intact classes at Georgetown University. There were 8 females and 10 males. Their mean age was 20.8 years and all except 3 participants were L1 English speakers. The non-English L1s were non-Romance languages: Bengali, Indonesian,
and Urdu. At the time of the study, participants had received approximately 30 hours of communicative classroom exposure to Spanish.

Since the goal of the pilot study was to test the materials, the two most contrasting experimental groups were used: unfocused direct WCF ($n = 6$) and unfocused indirect WCF ($n = 6$). Group one received unfocused direct WCF via crossing out and the provision of a correction above the error. Group two received unfocused indirect WCF via the underlining of errors. Group three served as the control group ($n = 6$) and participants completed the same task as the experimental groups, but they received commentary on the overall organization and content. Participants were instructed to TA during each phase of data collection. Table 14 below depicts the groups’ components:

Table 14

<table>
<thead>
<tr>
<th>Pilot Groups’ Components</th>
<th>$n$</th>
<th>WCF provision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfocused direct</td>
<td>6</td>
<td>cross out error + insert correction</td>
</tr>
<tr>
<td>Unfocused indirect</td>
<td>6</td>
<td>underline error</td>
</tr>
<tr>
<td>Control</td>
<td>6</td>
<td>feedback on content + overall organization</td>
</tr>
</tbody>
</table>

_Pilot linguistic targets._ Prior to the time of the pilot study, the participants had been exposed to the target forms in their classrooms. They had also explicitly practiced four of the five forms via homework assignments and classroom activities throughout the semester. Although they had not explicitly practiced the remaining form, they had been exposed to it during the semester. It is widely agreed-upon that L1 English speakers learning L2 Spanish have trouble with 1) the copular verbs (*ser* and *estar*) (i.e. Marin, 2004; Montrul & Slabakova, 2002;...
VanPatten, 2010), 2) gender agreement in adjectives (i.e. Montrul, Foote, & Perpiñán, 2008), and 3) past tense aspect (the preterit and the imperfect) (i.e. Rodríguez Prieto, 2009; Salaberry, 2008). With respect to *ser* and *estar*, Guntermann (1992) explained that “*ser* normally serves as an identifier, definer, or classifier of entities and events, *estar* to locate entities in space and to attribute to its subject a condition that is in contrast with what the speaker views as normal or expected qualities” (p. 1297). All of these concepts are expressed with one copula in English, “to be.” As for gender agreement, Sagarra and Herschensohn (2011) offer that L1 English learners of Spanish may have difficulty because “English has animate gender interpretation, but nouns like waiter/waitress bear no grammatical gender feature and denote a female in the same way that the proper name Alice conventionally does” (p. 169). As for aspect, García (2017) offers a possibility as to why L1 English learners of Spanish often struggle to select the appropriate past tense aspect. “Spanish uses the preterit to refer to bounded events with a clear beginning and/or end. However, while Spanish usually marks unbounded events with imperfect, English resorts to a wide variety of linguistic alternatives to convey the same notion” (p. 103).

In order to further scrutinize the copulas, they were broken down into categories of uses applicable to the study. Table 15 displays all of the targets, their distinct uses and forms that would potentially emerge in the data, and examples of each. The copula examples below appear in the present tense in order to highlight the uses rather than aspect. Additionally, the examples of the preterit and imperfect past tense conjugations include the first person singular, first person plural, third person singular, and third person plural forms because they are the most likely of all conjugations to arise in the compositions based on the prompt. Some of the examples are adaptations from the textbook, Vistazos, used in the Beginning Spanish classes in the participants’ courses.
Table 15

**Pilot Linguistic Targets**

<table>
<thead>
<tr>
<th>Linguistic Targets</th>
<th>Uses</th>
<th>Forms</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ser</em></td>
<td>a. Permanent characteristics</td>
<td>a. <em>Normalmente Logan es reservado.</em> Normally Logan is reserved.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. Location of event</td>
<td>b. <em>El concierto es en el Teatro Warner.</em> The concert is in the Warner Theater.</td>
<td></td>
</tr>
<tr>
<td><em>estar</em></td>
<td>a. Temporary changes in certain situations</td>
<td>a. <em>Hoy Logan está gregario.</em> Today Logan is gregarious.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. Location of people/objects</td>
<td>b. <em>Logan está en el Teatro Warner.</em> Logan is in the Warner Theater.</td>
<td></td>
</tr>
<tr>
<td>Gender agreement</td>
<td>Agreement with noun to which adjective refers, typically ending in <em>–o</em> or <em>–a</em></td>
<td>a. masc. <em>Logan</em> [masc.] <em>es alto</em> [masc.] Logan is tall.</td>
<td></td>
</tr>
<tr>
<td>Preterit</td>
<td>Completed actions, events, and states in the past</td>
<td>a. I SG <em>Yo fui</em> [I SG] <em>al concierto.</em> I went to the concert.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. I PL <em>Chris y yo fuimos</em> [I PL] <em>al concierto.</em> Chris and I went to the concert.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. III SG <em>Bobby fue</em> [III SG] <em>al concierto.</em> Bobby went to the concert.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d. III PL <em>Bobby y Marie fueron</em> [III PL] <em>al concierto.</em> Bobby and Marie went to the concert.</td>
<td></td>
</tr>
<tr>
<td>Imperfect</td>
<td>Actions, events, and states in progress in the past</td>
<td>a. I SG <em>Yo iba</em> [I SG] <em>a muchos conciertos.</em> I used to/would go to a lot of concerts.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>b. I PL <em>Chris y yo íbamos</em> [I PL] <em>a muchos conciertos.</em> Chris and I used to/would go to a lot of concerts.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>c. III SG <em>Bobby iba</em> [III SG] <em>a muchos conciertos.</em> Bobby used to/would go to a lot of concerts.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>d. III PL <em>Bobby y Marie iban</em> [III PL] <em>a muchos conciertos.</em> Bobby and Marie used to/would go to a lot of concerts.</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** The copular verbs *ser* and *estar* are shown in the present tense to emphasize the uses in the study, but participants composed in the past tense. The abbreviations are as follows: masc. for masculine, fem. for feminine, I for first person, III for third person, SG for singular, and PL for plural.
Below is an example of the appearance of WCF that the participants received per respective experimental group. The same example sentence is used for ease of comparison and the English translation is as follows: When I was a teenager, I went to a party.

\begin{verbatim}
era               fui
\end{verbatim}

Direct: Cuando estaba adolescente, iba a una fiesta.

Indirect: Cuando estaba adolescente, iba a una fiesta.

**Pilot materials.** A language background questionnaire was used in order to determine participants’ L1 and L2 and possible experience living in a Spanish-speaking country. The questionnaire was an adaptation of Lim and Godfroid (2014) and Calderón (2014), which is a modification of Freed, Dewey, Segalowitz, and Halter’s (2004) Language Contact Profile. (See Appendix A).

A TA practice activity, adapted from Calderón (2014), was used to ensure the participants understood how to TA during the tasks. The practice activity was altered to include Bowles’ (2010) suggestions for instructions to “include (1) a description of what is meant by ‘thinking aloud,’ (2) the language(s) participants are allowed to use to verbalize their thoughts, and (3) the level of detail and reflection required in the think-aloud” (p. 115). Macintosh computers were used in the language laboratory on campus at Georgetown University in order to utilize the audio and video recording program QuickTime Media Player. Participants composed their writing tasks on a word processing document in Microsoft Word. (See Appendix B for the TA practice activity).

For the sake of making the experimental task more pedagogical than those used in most previous studies (i.e. picture description tasks or story retellings) and for a more realistic, real-
life language use practice (i.e. Ellis, 2003; Stefanou & Révész, 2015), participants wrote a descriptive composition in which they were instructed to imagine they had been asked by a Spanish-language blog to post an entry about themselves and a friend of the opposite sex as teenagers in order to elicit agreement for the masculine and feminine genders in Spanish, the uses of _ser_ and _estar_, i.e. descriptions (_permanent characteristics versus temporary changes in certain situations_) and location (_event versus person or object_), and the preterit and imperfect, i.e. _actions completed in the past versus habitual actions in the past_. The prompt was provided in English to ensure the participants understood the topic. They were instructed to write 200-250 words. (See Appendices C for the prompt, which the participants received, and D for the prompt including the uses of the linguistic targets for clarification, which participants did not receive).

The researcher also placed a slip of paper containing accent mark typing instructions next to each keyboard in case participants were unfamiliar with typing on Macintosh computers. (See Appendix E).

**Pilot procedure.** Once the researcher obtained Institutional Review Board (IRB) approval for the pilot study (#2015-1288), she asked the Non-Intensive Spanish Language Program Director and Non-Intensive Beginning/Introductory Spanish Assistant Director to accommodate her for 3 days in the Beginning Spanish syllabus.

At Time 1 of the experiment, the researcher arrived at the language laboratory on campus before the participants. She placed a hardcopy of the IRB consent form and the TA practice task on each keyboard. She also set each Macintosh computer to Spanish mode, meaning the keyboards were able to produce the necessary diacritics and punctuation, and disabled autocorrect and all spelling and grammar check features on Microsoft Word. Participants and their instructors met the researcher in the laboratory during their normal class meeting times,
9am and 11am respectively. They sat at a computer of their choice, read the IRB consent form, and were given the opportunity to ask questions. Next, the researcher explained to the class how to TA and then the participants completed the TA practice task. They then put the IRB consent form and TA practice task aside, put on their headsets (headphones-microphone combination), clicked the record button on the QuickTime audio recording and screen recording programs, and the researcher distributed the composition prompt to them. They read the prompt and the researcher emphasized the instructions located at the bottom of the prompt that they would be permitted to use the online dictionary www.wordreference.com only to look up individual words. They were neither allowed to use the conjugation feature on this site nor consult any other dictionaries or sites for translations of phrases. TAs provided evidence of violations of this rule. Participants were given 30 minutes to complete the writing task. Each participant’s total time taken to complete the task was recorded and the researcher made a seating chart of the laboratory so the participants sat at the same computer for the remainder of the experiment.

After the Time 1 session, participants were randomly assigned to one of the three groups and then the researcher provided the respective WCF in green, a practice that mirrored the procedure used in the Spanish language classes that the participants had experienced on previous compositions and assignments prior to the experiment, and printed the compositions. This first draft served as a pretest for participants’ ability to accurately produce the linguistic targets, (ser, estar, adjective gender agreement, the preterit, and the imperfect).

Time 2 was two calendar days later and the researcher met the participants and their instructors again in the language laboratory at their normal class meeting times. The researcher arrived beforehand and placed facedown each participant’s printed composition with WCF on her/his corresponding keyboard from Time 1. She also saved each participant’s first draft to
her/his computer desktop, set the computer to Spanish mode, disabled autocorrect and spelling and grammar check on Microsoft Word, and readied the QuickTime recording program. As the participants entered, they completed the language background questionnaire. Once all participants had arrived and completed the questionnaire, they were instructed to flip over their composition hardcopies, open their first drafts from the desktop of their computers, revise their compositions based on the respective feedback provided, and TA. Participants had 15 minutes to complete their revisions and the researcher marked each participant’s total time on task. The second draft served as the posttest for *ser, estar*, adjective gender agreement, the preterit, and the imperfect.

Time 3 of the experiment was held one week after Time 1. Once again, the researcher arrived early and saved each participant’s first draft to her/his computer desktop, set the computer to Spanish mode, disabled autocorrect and spelling and grammar check on Microsoft Word, and readied the QuickTime recording program. She met the participants and their instructors in the language laboratory during their regular class meeting times. She instructed the participants to open their first drafts, without feedback, and asked them to revise their compositions to the best of their ability and to TA. They had 15 minutes to complete their revisions and the researcher noted each participant’s total time on task. The third draft functioned as the delayed posttest for *ser, estar*, adjective gender agreement, the preterit, and the imperfect. Figure 1 below represents a summary of the procedure.
Figure 1

**Summary of Pilot Data Collection Procedure**

| Time 1: 40 minutes | • IRB consent (5 minutes)  
|                    | • TA practice task (5 minutes)  
|                    | • Pretest: Draft 1, Prompt (30 minutes)  

2 days later

| Time 2: 20 minutes | • Language background questionnaire (5 minutes)  
|                    | • Immediate Posttest: Draft 2, Prompt, + WCF (15 minutes)  

5 days later

| Time 3: 15 minutes | • Delayed Posttest: Draft 3, Prompt, - WCF (15 minutes)  

**Pilot scoring and coding.** Written production accuracy of each linguistic target was calculated by the total number of instances of correct usage of each, divided by the total number of instances where each target structure was correctly used, incorrectly used or omitted from each draft. To borrow previous studies’ scoring procedure descriptions (Bitchener et al., 2005; Bitchener 2008; Bitchener & Knoch 2010b), if a participant displayed three correct uses of the target structure out of 10 obligatory instances in their text, then they earned a 30% accuracy score for the target structure on a particular draft. For example, Figure 2 shows the formula used to calculate accuracy, applied to the preterit.
This was repeated over the three drafts. The TAs allowed the researcher to hear the points at which each participant arrived while revising her/his drafts so she strictly included only the instances of target structures that appeared in all three drafts up to that point. In other words, if a participant wrote a total of 200 words at Time 1 but only revised the first 150 words at Time 2 then the researcher only analyzed the instances of target structures that appeared in the first 150 words of all three drafts.

In order to code for levels of DOP, the researcher adapted Leow’s (2015) coding scheme. Whereas Leow’s codes represent DOP of lexical items in L2 reading tasks, the modified codes in the pilot study represent DOP prompted by feedback on the copulas. Table 16 displays the adapted coding scheme for lexical items, which was adapted for semantic selection of *ser* or *estar* in the pilot study.
<table>
<thead>
<tr>
<th>Description</th>
<th>Level 1: Low DOP</th>
<th>Level 2: Medium DOP</th>
<th>Level 3: High DOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shows no potential for producing appropriate form with regard to intended meaning</td>
<td>Provides some evidence of processing the item</td>
<td>Provides evidence of making accurate form-meaning connection</td>
<td></td>
</tr>
<tr>
<td>Writes quickly, breezes over or does not produce target at all</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writes the target in English</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Says s/he does not know/forgets the word</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>If s/he looks it up, s/he does not spend much time processing it</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low level of cognitive effort to get the meaning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spends a bit more time processing target item (whether looking it up and spending time once found or producing it on her/his own)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Makes comments that indicate some processing of target item</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some level of cognitive effort to get meaning of target item</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spends time processing target item (whether looking it up and spending time once found or producing it on her/his own)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provides an accurate translation in English/Spanish</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High level of cognitive effort to get meaning of target item</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 17 shows the coding scheme of grammatical structures, which are adjective gender agreement following *ser* and *estar* and the use of the preterit and imperfect.

Table 17

*Pilot Coding Scheme for Depth of Processing of Feedback on Adjective Gender Agreement and Past Tense Aspect*

<table>
<thead>
<tr>
<th>Level 1: Low DOP</th>
<th>Level 2: Medium DOP</th>
<th>Level 3: High DOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Comments on target structure in relation to grammatical features</td>
<td>Arrives at an inaccurate, partially or fully accurate underlying grammatical rule</td>
</tr>
<tr>
<td>Descriptors</td>
<td>Spends a bit more time processing target structure</td>
<td>Makes hypotheses regarding target structure</td>
</tr>
<tr>
<td></td>
<td>Makes comments that indicate some processing of target structure, refers to a previous instance with the same form</td>
<td>Provides an inaccurate, accurate and/or partially accurate rule</td>
</tr>
<tr>
<td></td>
<td>Some level of cognitive effort to process target structure grammatically</td>
<td>Corrects previous translation</td>
</tr>
<tr>
<td></td>
<td>Spends much time processing target structure</td>
<td>Spends much time processing target structure</td>
</tr>
<tr>
<td></td>
<td>High level of cognitive effort to process target structure grammatically</td>
<td></td>
</tr>
</tbody>
</table>

The researcher only coded DOP on the linguistic targets the experimental group participants produced erroneously and on which they subsequently received WCF. She coded the control
participants’ DOP according to how they processed the linguistic targets they had produced incorrectly; on which they did not receive WCF but would have had they been assigned to an experimental group. The researcher and a colleague independently coded 25% of the TA transcriptions, met to discuss their results and found 99% agreement on both lexical and grammatical items. The researcher then coded the remaining 75% of the transcriptions.

In order to analyze further the participants’ DOP, the researcher calculated an average DOP score by dividing the number of instances of processing at one level of DOP into the total number of instances of processing at all three levels of DOP. Figure 3 below shows an example of this formula applied to the grammatical processing of the preterit.

Figure 3

Pilot Depth of Processing Score Formula

\[
\frac{\text{Total # instances of processing the preterit at a given DOP level}}{\text{Total # instances of processing the preterit at all DOP levels}} = \text{DOP score of the preterit}
\]

For example, if a participant showed evidence in her/his TA protocol of processing the preterit at a low level 10 times and processed the preterit at a medium level 1 time and at a high level 1 time, s/he received a DOP score of 83% LOW (10 divided by 12) for the preterit. The percentage was labeled HIGH, MEDIUM, or LOW based on the majority of the type of processing each participant applied.

If participants did not TA during any of the three data collection phases, they were eliminated from the analysis. Furthermore, if the researcher heard in the TA that participants used resources other than the permitted online dictionary, they were also eliminated from the
analysis. An original pool of 26 participants was filtered down to 18 after these exclusion criteria were applied.

**Pilot analysis.** The researcher ran statistical analyses in the Statistical Package for Social Sciences (SPSS) version 22 in the Gelardin New Media Center in Lauinger Library at Georgetown University. In order to address RQ1, accuracy scores and WCF types were submitted to a repeated measures two-way mixed ANOVA with a between-subjects and within-subjects design. However, only mean scores will be reported due to small cell sizes. As for RQ2, descriptive results were provided based on the coded TAs and ensuing DOP scores.

**Pilot results.** Participants were allotted 30 minutes to compose the first draft at Time 1, 15 minutes to revise with WCF if applicable at Time 2, and 15 minutes to revise without WCF at Time 3. At Time 1 participants’ average time on task was 32 minutes, 53 seconds. At Time 2 the entire group spent an average of 15 minutes, 56 seconds working on their revisions and at Time 3 participants averaged 14 minutes, 59 seconds revising.

With regard to the first research question in the pilot study, (Does type of unfocused written corrective feedback (direct or indirect) have a differential effect on adult L2 Beginning Spanish learners’ written production accuracy of the copula *ser*, the copula *estar*, adjective gender agreement, the preterit past tense aspect, and the imperfect past tense aspect? If so, does the effect on accuracy on each respective linguistic target last over 5 days?), participants’ accuracy scores for each linguistic target were submitted to repeated measures ANOVAs. It should be noted that while there were 18 participants in the pilot study, not every participant produced every linguistic target in the prompted context (i.e. Shintani et al., 2014). Therefore, the cell sizes differed for each linguistic target. Each cell size is listed in the descriptive statistics tables to follow. Furthermore, since the cell sizes were so small, only descriptive statistics are
presented. Also, the uses of *ser* (e.g., permanent characteristics and location) were collapsed into one category, as were the uses of *estar* (e.g., temporary change and location) since participants produced *ser* location 0.44 times per participant, *estar* location 0.56 times per participant, *estar* temporary change 0.89 times per participant, *ser* permanent description 3.44 times per participant, and even added an unprompted use of *estar*: condition, which they produced 0.44 times each. WCF type (direct, indirect, control) was the between-subjects factor and time (pretest, immediate posttest, delayed posttest) was the within-subjects factor. Tables 18, 19, 20, 21, and 22 below display the mean scores in percentages for the lexical item *ser*, the lexical item *estar*, adjective gender agreement, the preterit, and the imperfect.

Table 18

*Pilot Mean Scores for the Accurate Decision to Produce Ser*

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest Mean</th>
<th>Pretest SD</th>
<th>Immediate posttest Mean</th>
<th>Immediate posttest SD</th>
<th>Delayed posttest Mean</th>
<th>Delayed posttest SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>3</td>
<td>45.00</td>
<td>42.72</td>
<td>83.33</td>
<td>28.87</td>
<td>71.33</td>
<td>49.65</td>
</tr>
<tr>
<td>Indirect</td>
<td>3</td>
<td>75.00</td>
<td>43.30</td>
<td>53.33</td>
<td>50.33</td>
<td>65.00</td>
<td>40.93</td>
</tr>
<tr>
<td>Control</td>
<td>3</td>
<td>54.33</td>
<td>50.56</td>
<td>47.67</td>
<td>42.15</td>
<td>21.00</td>
<td>36.37</td>
</tr>
</tbody>
</table>
Table 19

*Pilot Mean Scores for the Accurate Decision to Produce Estar*

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest</th>
<th>Immediate posttest</th>
<th>Delayed posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Direct</td>
<td>2</td>
<td>41.50</td>
<td>12.02</td>
<td>100.00</td>
</tr>
<tr>
<td>Indirect</td>
<td>2</td>
<td>50.00</td>
<td>70.71</td>
<td>100.00</td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
<td>8.33</td>
<td>20.41</td>
<td>25.00</td>
</tr>
</tbody>
</table>

Table 20

*Pilot Mean Scores for the Accurate Production of Adjective Gender Agreement*

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest</th>
<th>Immediate posttest</th>
<th>Delayed posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>Direct</td>
<td>2</td>
<td>68.00</td>
<td>16.97</td>
<td>100.00</td>
</tr>
<tr>
<td>Indirect</td>
<td>2</td>
<td>51.50</td>
<td>12.02</td>
<td>62.00</td>
</tr>
<tr>
<td>Control</td>
<td>5</td>
<td>57.80</td>
<td>25.72</td>
<td>58.60</td>
</tr>
</tbody>
</table>
### Table 21

**Pilot Mean Scores for the Accurate Production of the Preterit**

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest Mean</th>
<th>Pretest SD</th>
<th>Immediate posttest Mean</th>
<th>Immediate posttest SD</th>
<th>Delayed posttest Mean</th>
<th>Delayed posttest SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>6</td>
<td>22.50</td>
<td>13.22</td>
<td>79.33</td>
<td>27.19</td>
<td>32.83</td>
<td>24.00</td>
</tr>
<tr>
<td>Indirect</td>
<td>6</td>
<td>26.50</td>
<td>15.60</td>
<td>27.67</td>
<td>12.74</td>
<td>30.83</td>
<td>18.71</td>
</tr>
<tr>
<td>Control</td>
<td>6</td>
<td>28.50</td>
<td>15.11</td>
<td>28.00</td>
<td>15.48</td>
<td>26.00</td>
<td>16.85</td>
</tr>
</tbody>
</table>

### Table 22

**Pilot Mean Scores for the Accurate Production of the Imperfect**

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>Pretest Mean</th>
<th>Pretest SD</th>
<th>Immediate posttest Mean</th>
<th>Immediate posttest SD</th>
<th>Delayed posttest Mean</th>
<th>Delayed posttest SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>6</td>
<td>0.00</td>
<td>0.00</td>
<td>82.50</td>
<td>32.39</td>
<td>60.17</td>
<td>30.93</td>
</tr>
<tr>
<td>Indirect</td>
<td>6</td>
<td>19.67</td>
<td>22.47</td>
<td>20.67</td>
<td>23.95</td>
<td>40.67</td>
<td>31.00</td>
</tr>
<tr>
<td>Control</td>
<td>6</td>
<td>15.17</td>
<td>31.44</td>
<td>16.50</td>
<td>31.29</td>
<td>25.00</td>
<td>31.50</td>
</tr>
</tbody>
</table>

At the time of the pretest, the direct WCF, indirect WCF, and control groups’ mean scores were not comparable so the scores will be presented by individual group. At the immediate posttest, the direct WCF group’s mean scores spiked and then declined at the delayed posttest and, in the case of adjective gender agreement, the score worsened compared to the pretest, but increased for all other linguistic targets. As for the indirect WCF group, the mean
scores increased at the time of the immediate posttest except for *ser*. They subsequently increased for *ser*, the preterit, and the imperfect at the time of the delayed posttest and decreased for *estar* and adjective gender agreement. In the case of *ser*, the delayed posttest score was lower than the pretest score. All other linguistic items’ mean scores improved from the pretest to the delayed posttest. The control groups’ scores all increased from the pretest to the immediate posttest except for the preterit score, which decreased slightly. The mean scores for *ser*, adjective gender agreement, and the preterit decreased from the immediate to the delayed posttest, but the mean scores for *estar* and the imperfect increased. The scores for *ser*, adjective gender agreement, and the preterit also decreased from the pretest to the delayed posttest and increased once again for *estar* and the imperfect.

With regard to RQ2, (How do adult L2 Beginning Spanish learners process unfocused written corrective feedback?), the researcher noted the total number of instances of processing for lexical items (i.e. the decision to produce *ser* or *estar* in a given context) and the total number of instances of processing for grammatical structures (i.e. agreement or the decision to use the preterit or imperfect in a given context). As evidenced by the TA protocols, most participants processed the feedback with respect to the target structures grammatically. Specifically, the researcher heard a total of 14 instances of lexical processing and 257 instances of grammatical processing. The imbalance in type of processing was not unexpected since the prompt elicited more target grammatical structures than lexical items, leaving more risk of errors and subsequently more WCF for the former.

The researcher calculated the DOP scores per linguistic target and found that the direct group processed *ser* at a medium level the majority of the time, *estar* at a high level every time, adjective gender agreement at a high level the majority of the time, the preterit at medium and
low levels half the time, each, and the imperfect at a low level most of the time. The indirect
group processed ser at a medium level the majority of the time, estar at a high level every time,
adjective gender agreement at a high level every time, the preterit at a low level half the time,
and the imperfect at a low level about half the time as well. The control group processed all
linguistic targets at a low level. Table 23 below provides a summary of DOP scores by group.

Table 23

*Pilot Depth of Processing Scores*

<table>
<thead>
<tr>
<th></th>
<th>ser</th>
<th>estar</th>
<th>Adjective gender agreement</th>
<th>Preterit</th>
<th>Imperfect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>66%</td>
<td>100%</td>
<td>67%</td>
<td>50% MEDIUM/50%</td>
<td>55% LOW</td>
</tr>
<tr>
<td></td>
<td>MEDIUM</td>
<td>HIGH</td>
<td>HIGH</td>
<td>50% LOW</td>
<td></td>
</tr>
<tr>
<td>Indirect</td>
<td>66%</td>
<td>100%</td>
<td>100%</td>
<td>50% LOW</td>
<td>49% LOW</td>
</tr>
<tr>
<td></td>
<td>MEDIUM</td>
<td>HIGH</td>
<td>LOW</td>
<td>LOW</td>
<td>LOW</td>
</tr>
<tr>
<td>Control</td>
<td>83%</td>
<td>100%</td>
<td>100%</td>
<td>100% LOW</td>
<td>100% LOW</td>
</tr>
<tr>
<td></td>
<td>LOW</td>
<td>LOW</td>
<td>LOW</td>
<td>LOW</td>
<td>LOW</td>
</tr>
</tbody>
</table>

Below are some excerpts from the TA protocols to exemplify participants’ processing.
Translations are provided in brackets where necessary. Capitalized letters within words represent
participants’ emphasis.

High DOP:

“Um divertida [fun, feminine] ‘cause it’s a party which is feminine.” (adjective gender
agreement, direct WCF)
“I USED to have a lot of- USED TO being imperfect but I don’t know anything in the imperfect except I know it goes -ba -bas -bamos -ban but I don’t know what kind of stem. I guess I could technically piece it together. Estu- Hm.” (imperfect, indirect WCF)

Medium DOP:

“…fiesta fue- [party was] ah I get it. Because it has to be ser not estar.” (ser, direct WCF)

“Mi favorita [My favorite]. La clase de biología...mi favorita [Biology class, feminine...my favorite]. Favorita [Favorite]. Why is my concordancia [agreement] terrible [pronounced in Spanish]?” (adjective gender agreement, indirect WCF)

Low DOP:

“I attended. Asi- Yo asistÍ [I attended].” (preterit, direct WCF)

“Estuve [I was]. Changing all the fui’s [form of ser] to estuve [form of estar].” (estar, indirect WCF)

“Fui una estudiante [I was a student].” (imperfect, control)

In addition to semantic and grammatical processing, participants made comments about the respective WCF they received. Below are a few examples heard in the TA protocols.

Direct: “I have no idea where any of these edits are coming from. I mean some of them are obvious ones like not using an accent on más [more] or forgetting to do A mi amiga le gusta [My friend likes] and después DE [after]. I get why that makes sense but for any of the verb forms…”

Indirect: “Uh nothing is coming to me um based on these edits. I mean I’m changing them for the sake of changing them because they’re underlined. Well some of them. Others I just have absolutely no idea.”
Control: “Not quite sure what else to do because I think the scope of my knowledge has been maximized here.”

**Pilot discussion.** The descriptive results of the pilot study must be interpreted with caution due to the low cell sizes for the respective linguistic targets. Overall, the direct group displayed a pattern of increase-decrease. At the immediate posttest scores increased and at the delayed posttest scores decreased. This pattern was not surprising because the corrections were given to the participants at the time of the immediate posttest and made unavailable at the delayed posttest. The unexpected finding, though, was that they processed the “spoon-fed” corrections at overall medium and high levels, except in the case of the imperfect. The researcher was surprised because she had anticipated processing at low and medium levels since the participants did not have to solve problems but rather could have copied the corrections. The Hawthorne effect could have been at play here, since the participants were taking part in an experiment coupled with completing the assignment for a grade. As for the indirect group, a few varied patterns surged. The first was an increase at the immediate posttest and decrease at the delayed posttest for *estar* and adjective gender agreement, the second was an increase at the immediate posttest and another increase at the delayed posttest for the preterit and imperfect, and the last was not a pattern but rather a solitary case of a decrease at the immediate posttest and slight increase, but still lower than the pretest, at the delayed posttest for *ser*. The increase in the imperfect score over time was not a shock since the participants explicitly practiced this form in class between the immediate posttest and the delayed posttest. As for the preterit, the same reason could be applied since the participants were beginning to juxtapose the two forms in class. With regard to *estar* and adjective gender agreement, although the scores followed a similar pattern over time, participants’ scores nearly returned to the pretest score when they processed
estar at a high level and they stayed near the immediate posttest level at the delayed posttest for adjective gender agreement despite processing at a low level. However, the overall means were similar over time. As for ser, the TA protocols allowed the researcher to hear some participants who had erroneously produced ser and subsequently received an underline simply change all ser occurrences to estar. However, they processed at a medium level overall and their scores never returned to the pretest score. For the control group, despite only receiving feedback on general content and overall organization, the participants self-edited and even showed patterns of improvement for estar and the imperfect despite processing these and the other linguistic targets at a low level. The uptick in the accurate production of the imperfect can be explained the same way it was for the indirect group: the participants explicitly practiced this form in class between the immediate and delayed posttests.

_Pilot methodological issues and modifications._ With regard to timing of data collection, as mentioned above, the pilot study occurred during a time in the semester when the participants had already explicitly practiced four of the five target structures. The decision to collect data at this time was made based on the course syllabus and the three days chosen were the only available days since the participants would be undergoing oral exams and final exam review after they had explicitly practiced the final linguistic target. As such, they studied this target at home and practiced it in class between the immediate and delayed posttest. Although the results were not significantly affected by this practice, it was a methodological risk and was adjusted for the larger study. The Spanish Language Program Director and the Assistant Director of Beginning & Introductory Spanish approved an alteration in the syllabus for the Fall 2016 semester so that all linguistic targets involved in the study would be presented and practiced before the start of the current study.
Secondly, the researcher had planned to use the rubric from the Spanish language program to rate the quality of the first draft but when it was discovered that only six out of 18 participants met the minimum word count requirement on the first draft, the researcher consulted with the Spanish Language Program Director and they decided to offer the participants the option to finish their compositions at home after the pilot study ended the following week. Therefore, the researcher did not provide the participants with the rubric or their grades until after the experiment had ended. Their compositions were graded because they counted as the fourth and final composition of the course. Additionally, one of the instructors opted to grade his section’s compositions. The researcher graded the other instructor’s section’s compositions as originally planned. (See Appendix F for the email script to instructors, which includes the consent form students read at Time 1).

In the main study, data were collected from volunteer participants from Beginning and Introductory I Spanish courses outside of class time. The composition was not a course requirement and did not count toward a grade. The prompts for the Beginning and Introductory Spanish course compositions were designed to provide students practice with grammatical points, other than those in the study, covered in class throughout the semester. The main study composition prompt provided the opportunity to practice the grammar presented over the course of the few weeks leading up to the time of the study so it was ecologically valid in that participants had an opportunity to 1) practice the vocabulary and grammar they had been working with in class and 2) complete a potential real-world task. Since data were collected outside of class, participants had more time to meet the minimum word count requirement. (See Appendix G for the email recruitment script to participants).
Thirdly, some participants did not produce every linguistic target in their compositions. For example, one participant answered with action verbs to describe himself and his friend instead of using *ser* and *estar* with adjectives. Furthermore, 8 participants used the words *introvertido/-a* (introverted) and *extrovertido/-a* (extroverted), which pair with *ser* and not with *estar*, to describe themselves and their friends. The researcher decided to exclude all cases of *ser* and *estar* when they appeared with *intro-/extrovertido/-a* (intro-/extroverted) after she consulted the course textbook, and two native Spanish speakers who did not accept *estar + intro-/extrovertido/-a* (to be intro-/extroverted) even in the context of temporary change. Since *introvertido/-a* (introverted) and *extrovertido/-a* (extroverted) appear on a vocabulary list in the textbook in a lesson on *ser* and *estar*, they could not be avoided in the current study and all instances of these specific copula + adjective pairings that appeared in participants’ compositions were eliminated from analysis.

Moreover, the participants did not produce an overall sufficient number of items of *ser* or *estar* per use for statistical analysis. While the prompt was designed to elicit at least 2 (location), 4 (temporary change), and 6 (permanent description) uses in the first draft, respectively, participants produced *ser* location 0.44 times per participant, *estar* location 0.56 times per participant, *estar* temporary change 0.89 times per participant, *ser* permanent description 3.44 times per participant, and even added an unprompted use of *estar: condition*, for which they produced 0.44 times each. The prompt was updated for the main study in order to more tightly control for ample productions of the discrete uses. Namely, the main study prompt was designed to elicit at least 8 uses of *ser* permanent description, 8 uses of *estar* temporary change, 8 uses of *ser* location, and 8 uses of *estar* location. As for adjective gender agreement, the prompt was designed to elicit at least 4 occurrences of masculine agreement and 4 occurrences of feminine
agreement, for a total of 8 uses of adjective gender agreement. Since the preterit and imperfect are embedded throughout the prompt, which consists of open-ended questions, the number of instances of these targets was not controlled. (See Appendices H and I for the main study prompt and the prompt with uses).

The obligatory occasion analysis was going to be utilized in the current study, as it was in the pilot, to determine the scores of each target, but the researcher decided it would be of more pedagogical interest to probe into the probable issues participants would have with the decision between *ser* and *estar* (i.e. Marín, 2004; Montrul & Slabakova, 2002), and the preterit and imperfect (i.e. Rodríguez Prieto, 2009; Salaberry, 2008). Therefore, the written production accuracy scores were calculated in a more discerning manner than in the pilot study to focus exclusively on the choice between using *ser* or *estar* and the preterit or imperfect. The obligatory occasion analysis, then, was not used since it is too vague to address dichotomous choices. The TA instances were transcribed and cleansed to include only instances of participants’ verbalizations regarding the decision between copulas or past tense aspects. For example, in some cases, participants left verbs out of sentences altogether, e.g. *Mi entrevista en un café* [My interview in a café]. In the obligatory occasion analysis, this error would have been classified as an omission but in the current analysis this type of error was not included because it does not represent the choice between copulas or aspects. On the other hand, if a participant produced, *Mi entrevista estuvo en un café* [My interview was- error- in a café], where *estuvo* should have been *fue*, the error was included in the analysis because it was an erroneous production in the *ser* versus *estar* dichotomy.

Another scoring change for the current study is instead of taking the overall accuracy score of the targets (as was done in the pilot study), the researcher only counted the targets that
appeared across all three drafts in the written production accuracy score for each target. To continue the example from above, if a participant wrote *Mi entrevista estuvo en un café* in Draft 1 and deleted the entire sentence from Draft 2, then it was not included in the score. Likewise, if the participant kept the sentence but changed *estuvo* to *ocurrió* [happened] then it was not included because *ocurrir* [to happen] is not *ser* or *estar*.

The use of technology did not go smoothly in the pilot study. Screen recordings were only taken during the first session of Time 1 because, presumably, the files were so large they froze the computers, which resulted in the loss of 63% of the screen recordings and participants were inconvenienced with having to change computers mid-composition. A few extra moments were allowed after the interruption to recuperate total time on task, which still averaged almost 33 minutes instead of the originally planned 30 minutes. The researcher spoke with the language laboratory staff and they suggested using a newly installed screen-capture and audio recording program called Echo360.

In terms of experimental groups, one more group was added to the current study: the unfocused metalinguistic WCF group. This feedback comes in the form of abbreviations and symbols, accompanied by a key. The symbols key provided to the unfocused metalinguistic WCF participants came from the Non-Intensive Spanish Program’s course materials and the participants were familiar with it since they had received this type of WCF on three compositions by the time of the study. (See Appendix J for the metalinguistic WCF abbreviations and symbols key).

**Current Study**

The above-discussed adjustments were applied to the current study. It was administered in Weeks 11, 12, and 14 of the Fall 2016 semester.
Methodology.

Participants. Participants in the current study were 61 students from 7 sections of Non-Intensive Beginning Spanish (SPAN 001) and Introductory I Spanish (SPAN 003), and 2 sections of Intensive Basic Spanish (SPAN 011) at a Mid-Atlantic university in the United States. There were 35 females and 26 males. Their mean age was 19.08 years and all except 9 participants were L1 English speakers. Of these 9 participants, 3 participants noted that English and another language were both their L1s: English/Cantonese, English/Chinese, and English/Vietnamese. The non-English L1s were non-Romance languages: Amharic \(n = 2\), Hindi/Punjabi \(n = 1\), German \(n = 1\), Korean \(n = 1\), and Vietnamese \(n = 1\). At the time of the study, participants had received approximately 28 hours of communicative classroom exposure to Spanish at the University. They had written 3, 2-draft compositions as per their course assignment requirements by the time of the study.

The experimental groups were as follows: unfocused direct WCF \(n = 15\), unfocused metalinguistic WCF \(n = 16\), unfocused indirect WCF \(n = 15\). Group one received unfocused direct WCF via crossing out and the provision of a correction above the error. Group two received feedback in the form of metalinguistic abbreviations accompanied by a sheet to help them decipher the abbreviations and see examples. Group three received unfocused indirect WCF via the underlining of errors. Group four served as the control group \(n = 15\) and participants completed the same task as the experimental groups, but they received commentary on the overall organization and content (i.e. Bitchener, 2008) in lieu of WCF on linguistic errors. All participants were instructed to TA during all phases of data collection. Table 24 below depicts the groups’ components.
### Table 2

**Groups’ Components**

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>WCF provision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unfocused direct</td>
<td>15</td>
<td>cross out error + insert correction</td>
</tr>
<tr>
<td>Unfocused metalinguistic</td>
<td>16</td>
<td>metalinguistic abbreviation codes + accompanying sheet with examples</td>
</tr>
<tr>
<td>Unfocused indirect</td>
<td>15</td>
<td>underline error</td>
</tr>
<tr>
<td>Control</td>
<td>15</td>
<td>feedback on content + overall organization</td>
</tr>
</tbody>
</table>

**Target dichotomies.** Bitchener (2008) and Ellis et al. (2008) highlighted the importance of selecting a target form that has been “partially learned” (Bitchener, 2008, p. 115) or “partially acquired” (Ellis et al., 2008, p. 357) since it will likely appear in the participants’ writing. In other words, participants can only produce the target items and structures if they have some knowledge of them. Prior to the time of the current study, participants had been exposed to the targets in their classrooms. They had also explicitly practiced them via homework assignments and classroom activities throughout the semester.

The same targets from the pilot study were used in the current study with a few changes. First, gender agreement in adjectives was omitted from the analysis. It was discovered after data collection that participants produced few to no tokens of gender agreement either by not following instructions to write about themselves and a friend of the opposite sex in their first drafts, by scoring at ceiling, or by producing adjectives that did not end in –o or –a. Second, as seen in the previous chapter in the pilot methodological issues and modifications section, in order to investigate learners’ probable issue with choosing between the copulas *ser* and *estar* (i.e.
Marín, 2004; Montrul & Slabakova, 2002) or the past tense aspects the preterit and the imperfect (i.e. Rodríguez Prieto, 2009; Salaberry, 2008), they were analyzed as dichotomies rather than individual targets. These target dichotomies were in turn *ser* versus *estar* and the preterit versus the imperfect. As in the pilot, the uses of each copula were broken down into discrete categories (*ser*: permanent descriptions or location of events; *estar*: temporary changes in certain situations or location of people or objects) and elicited as such via the prompt.

Below is an example of the appearance of WCF that the participants received per respective experimental group. The same example sentence is used for ease of comparison and the English translation is as follows: I went to a concert with my family and I was very excited that day.

Direct: *Iba al concierto con mi familia y era muy emocionada ese día.*

Metalinguistic: *Iba al concierto con mi familia y era muy emocionada ese día.*

Indirect: *Iba al concierto con mi familia y era muy emocionada ese día.*

**Materials.** The same language background questionnaire from the pilot study was used in the current study in order to determine participants’ L1, L2, and in some cases L3, and possible experience traveling to or living in a Spanish-speaking country. (See Appendix A).

The same TA practice activity from the pilot study, adapted from Calderón (2014), was used to ensure the participants understood how to TA during the tasks. (Again, see Appendix B for the TA practice activity).
Macintosh computers were used in the language laboratory on campus at the participants’ university in order to utilize the platform Echo360 for simultaneous audio and video recordings. Participants composed their writing tasks on a word processing document in Microsoft Word.

The prompt provided participants with contextual instructions, meaning they were going to write a blog entry for a Spanish-language site, and the questions they were instructed to address appeared in the form of a checklist. They were instructed to check each box as they composted their descriptive composition (blog entry) so they were sure to complete the task and so their compositions were as comparable as possible. As in the pilot, the prompt was provided in English to ensure the participants understood the topic and questions. They were instructed to write 200-250 words. (See Appendices C for the prompt, which the participants received, and D for the prompt including the uses of the targets for clarification, which participants did not receive).

Accent mark typing instructions were placed next to each keyboard in case participants were unfamiliar with typing on Macintosh computers. (See Appendix E).

The metalinguistic WCF participants used the symbols key during Time 2. (See Appendix J for the metalinguistic WCF abbreviations and symbols key).

Lastly, the participants responded to a small set of debriefing questions during Time 3 (the final session) of data collection. (See Appendix K).

**Procedure.** First, approval for the current study (#2016-0758) was obtained from the Institutional Review Board (IRB).

At Time 1 of the study, the researcher arrived at the language laboratory prior to the participants’ scheduled appointments to prepare their desktops. As in the pilot, she placed a hardcopy of the IRB consent form and the TA practice task on each keyboard. She also set each
Macintosh computer to Spanish mode and disabled autocorrect and all spelling and grammar check features on Microsoft Word. Additionally, a window in Google Chrome was open to www.wordreference.com.

Participants met the researcher in the laboratory at their scheduled times. They sat at a computer of their choice, read the IRB consent form, and were given the opportunity to ask questions. Next, the researcher explained how to TA and then the participants completed the TA practice task. They then put the IRB consent form and TA practice task aside, put on their headsets, clicked the record button on the Echo360 audio and video recording program, and the researcher distributed the composition prompt to them. They read the prompt and the researcher emphasized that they would be permitted to use the online dictionary on www.wordreference.com only to look up individual words and not to use the conjugation feature for verbs. The audio and/or video recordings provided evidence of violations of this stipulation. Participants were given as much time as they needed to complete the writing task. Each participant’s total time spent to complete the task was recorded on Echo360. They spent an average of 54 minutes, 46 seconds composing the first draft. The average word count was 338.6 words per composition.

After the Time 1 session, participants were randomly assigned to one of the four groups and then provided the respective WCF. This first draft served as a pretest for participants’ ability to accurately produce the target dichotomies, (ser versus estar and the preterit versus the imperfect).

Time 2 occurred one week later and the researcher met the participants again in the language laboratory. The researcher arrived beforehand and placed facedown each participant’s printed composition with WCF on her/his corresponding keyboard from Time 1. She also saved
each participant’s first draft to a computer desktop, set the computer to Spanish mode, disabled autocorrect and spelling and grammar check on Microsoft Word, readied the Echo360 recording program, and opened www.wordreference.com on Google Chrome. As the participants entered, they sat at the computer the researcher indicated to each of them via post-it notes at their seats and completed the language background questionnaire. Once all participants had arrived at their respective scheduled times and completed the questionnaire, they were instructed to flip over their composition hardcopies, open their first drafts on their computers, revise their compositions based on the respective feedback provided, and TA. Participants had unlimited time to complete their revisions and their total time on task was recorded on Echo360. They spent an average of 30 minutes, 55 seconds revising the draft. The metalinguistic WCF group took the longest amount of time, arguably because they a) had to decipher the symbols and codes and b) it was discovered that six of the sixteen participants had never seen nor used the symbols key. Table 25 displays time on task by group.

Table 25

*Average Time on Task per Written Corrective Feedback Type at Time 2*

<table>
<thead>
<tr>
<th>WCF Type</th>
<th>Average (min:sec)</th>
<th>Minimum (min:sec)</th>
<th>Maximum (min:sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>21:20</td>
<td>16:05</td>
<td>46:41</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>45:51</td>
<td>22:00</td>
<td>89:52</td>
</tr>
<tr>
<td>Indirect</td>
<td>34:02</td>
<td>9:19</td>
<td>59:25</td>
</tr>
<tr>
<td>Control</td>
<td>22:28</td>
<td>14:19</td>
<td>37:19</td>
</tr>
</tbody>
</table>

The second draft served as the immediate posttest for the target dichotomies. Figure 4 shows an example of a participant’s computer screen during Time 2. The cursor appears in red in Echo360
so the viewers can follow the user’s location. The participant has highlighted the verb *estuvo*, and in the next frame (not shown here) he changed *estuvo* to *fue*.

Figure 4

*Example of a Participant’s Computer Screen during Time 2*

![Example of a Participant’s Computer Screen during Time 2](image)

Time 3 of the experiment was held three weeks after Time 1. Once again, the researcher arrived early and saved each participant’s first draft to a computer desktop, set the computer to Spanish mode, disabled autocorrect and spelling and grammar check on Microsoft Word, and readied the Echo360 recording program and opened www.wordreference.com on Google Chrome. She met the participants in the language laboratory at their scheduled times. She instructed the participants to open their first drafts and asked them to revise their compositions to the best of their ability without any feedback and to TA. The audio and/or video recordings provided evidence of violations of this stipulation. They had unlimited time to complete their revisions and the researcher noted each participant’s total time on task, provided by Echo360. They spent an average of 18 minutes, 35 seconds on the third draft. Once they finished their final
revisions to Draft 3, they answered a set of debriefing questions about their experiences and opinions regarding the feedback they received. The third draft functioned as the delayed posttest for the written production accuracy scores of the target dichotomies. Figure 5 below represents a summary of the procedure and materials provided in each session of the current study.

Figure 5

Summary of Data Collection Procedure

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>• IRB consent</td>
<td>• Language background questionnaire</td>
<td>• Delayed Posttest: Draft 3, Prompt , − WCF</td>
</tr>
<tr>
<td>• TA practice task</td>
<td>• Immediate Posttest: Draft 2, Prompt, + WCF</td>
<td>• Debriefing questions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Pretest: Draft 1, Prompt</td>
<td>• Language background questionnaire</td>
<td>• Delayed Posttest: Draft 3, Prompt , − WCF</td>
</tr>
<tr>
<td>• IRB consent</td>
<td>• Immediate Posttest: Draft 2, Prompt, + WCF</td>
<td>• Debriefing questions</td>
</tr>
<tr>
<td>• TA practice task</td>
<td>• Immediate Posttest: Draft 2, Prompt, + WCF</td>
<td>• Debriefing questions</td>
</tr>
</tbody>
</table>

Scoring and coding. The written production accuracy score calculation was made by taking the total number of instances of correct usage of each, divided by the total number of instances where each target was correctly or incorrectly used. “Incorrect” refers to the erroneous choice between ser and estar or the preterit and imperfect. In this calculation, only the correct or incorrect choice between the targets is reflected. The copulas were deemed correct or incorrect regardless of verb tense/aspect so as to only score the choice between ser and estar and not to confute the analysis with verb aspect. In some cases, participants produced the copulas in the present tense. At first it would seem as though the feedback was confounded, as it was indicating
to them that they needed to change both the copula and the tense, but since the feedback was indeed to change tense and not aspect, these instances were included in the copula score and not the aspect score. The aspects were counted as correct or incorrect only when participants chose to produce one over the other. The copulas were excluded from this part of the analysis to avoid conflation of the two target dichotomies. Lastly, instances of misspelling, erroneous accent mark addition or omission, and incorrect subject-verb agreement were disregarded. For example, *Mis amigos fuimos conmigo al concierto* [My friends went –I PL– with me to the concert] was included as a correct production of the preterit, regardless of the erroneous subject-verb agreement.

Participants produced certain adjectives following the copulas and verbs in the past tense that were deemed beyond their level at the time of the study. While the researcher provided the respective WCF, these cases were disqualified from the analyses. As with the pilot study, if a participant wrote *ser* or *estar* extrovertido/a, the researcher provided corrective feedback but did not count this token in the *ser* versus *estar* analysis, as this adjective is exclusively paired with *ser* regardless of context. As for the preterit versus imperfect analysis, the following verbs were excluded: *conocer* (to meet in the preterit versus to know in the imperfect), *saber* (to find out in the preterit versus to know in the imperfect), *querer* (to intend in the preterit versus to want in the imperfect).

The researcher checked her own provision of WCF by marking a random sample of 25% of the compositions twice (i.e. Polio, 1997; Polio & Shea, 2014) and the intra-rater agreement came to 98.56%.

In order to code for levels of DOP in the main study, the researcher further adapted Leow’s (2015) coding scheme beyond the changes made for the pilot study. The modified codes
in the current study represent DOP of the target dichotomies: *ser* versus *estar* and the preterit versus imperfect, as prompted by the WCF. Table 26 displays the adapted coding scheme.

**Table 26**

*Coding Scheme for Depth of Processing of Ser versus Estar and the Preterit versus Imperfect*

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

DOP was only coded on the targets the experimental group participants produced erroneously and on which they subsequently received WCF. Control participants’ DOP was coded according
to how they processed the targets they had produced incorrectly; on which they did not receive WCF but would have had they been assigned to an experimental group. The researcher and a colleague independently coded a random sample of 25% of the TA transcriptions. Interrater agreement was calculated by dividing the total number of independently-coded and agreed-upon instances (between the two raters) into the total number of coded instances. The raters arrived independently at 97.1% agreement, discussed their discrepancies, and came to 100% agreement. The researcher then coded the remaining 75% of the transcriptions.

The DOP scoring system in the pilot study captured frequency of processing instances and not necessarily DOP itself. In order to analyze further the participants’ DOP in terms of its potential relationship to participants’ production scores on the target dichotomies, Zamora’s (2017) DOP scoring system was adopted. An average DOP score was calculated by assigning the following numeric scores to each TA instance (regarding *ser* versus *estar* or the preterit versus imperfect): instances of low levels of processing were given a 1, instances of medium levels were given a 3, and instances of high levels were given a 5.

Once each instance had a numeric value, all the values from each participant’s TA protocol were added together and the sum was divided by the total number of instances in the protocol. The quotient represented an average DOP score. A range of average DOP scores was defined and the corresponding levels of DOP were assigned. These corresponding levels of DOP represented each participant’s overall DOP for each target dichotomy. In other words, each participant’s TA was coded twice: once for DOP of *ser* versus *estar* and once for DOP of the preterit versus imperfect. As described above, instances of processing for two errors in one production, i.e. *ser* appeared in the preterit when it should have been *estar* in the imperfect, were eliminated from further analysis as the targets were intertwined. Since the average DOP score
ranges are quite arbitrary, some numeric distance was added to act as a safeguard between levels of processing. DOP scores that did not fall within the ranges were eliminated. DOP score ranges and their corresponding DOP levels are presented in Table 27.

Table 27

*Depth of Processing Scores*

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

As for running correlations between DOP score and accuracy score, they were going to be calculated by submitting the Draft 2 and Draft 3 scores, separately, with the DOP scores for each target dichotomy. However, in most pre-post-delayed posttest designs the treatment occurs between the pretest and immediate posttest and in the current study, the treatment (WCF) was administered during the immediate posttest (Draft 2) itself. There would have been no way to determine if the DOP scores correlated with the Draft 2 accuracy scores due to the simultaneous nature of the treatment and assessment. Likewise, Draft 1 served as the pretest but the participants did not start at 0 or even a below-chance score. Therefore, if correlations were run for Draft 2, the results would not reflect processing and what actually happened with respect to their accuracy scores between Draft 1 and Draft 2.

Participants were eliminated from the analyses for the following reasons: 1) if they did not follow instructions during the three data collection phases (i.e. if they used dictionaries other
than www.wordreference.com), 2) if they scored 65% or higher when making the choice to produce *ser* versus *estar* in the first draft, or 3) if they scored 90% or above when making the choice to produce the preterit versus imperfect on the first draft. The cut-off points for the copulas and verb aspects were set so high because most participants performed considerably well on the first draft, especially on the preterit versus imperfect. Originally, the researcher sought to assign the cut-off points at a below-chance score, but when it became apparent that the participants in the current study performed considerably well on Draft 1 with regard to the target dichotomies, it was decided these high cut-off points, while not desirable, were acceptable based on a few previous WCF studies, (e.g. Ellis et al., 2008; Sheen et al., 2009). In these studies, if participants scored 90% or higher, they were excluded from the data. Ellis et al. and Sheen et al. cited Brown (1973), who argued that “90% accuracy is a common cut-off point for grammatical features that learners have indeed learned.

The original pool of 61 participants was filtered down to 58 after it was discovered through the TAs and/or Echo360 video recordings that 2 participants did not follow instructions and 1 participant stopped her Echo360 recording 17 seconds into the session. After the target dichotomies were scored and cleansed by the respective cut-off points and confounding factors (i.e. the copulas were in the past tense and had subsequent errors for each dichotomy—*ser* appeared in the preterit when it should have been *estar* in the imperfect), 40 participants remained for the analysis.
Chapter IV: Results

This chapter presents qualitative and quantitative results to answer the three research questions.

Research Questions

The first research question was qualitative in nature. Therefore, the results will be presented as such. The second and third research questions were quantitative and the results will be presented accordingly.

Research question one. Research question one (RQ1) addressed how adult L2 Beginning Spanish learners process unfocused written corrective feedback on *ser* versus *estar* (RQ1a) and the preterit versus imperfect past tense aspects (RQ1b). The results are shown via excerpts from the TA protocols to exemplify participants’ DOP. As a reminder from the DOP scoring system, a TA instance that qualified as an example of high DOP does not imply that participants were classified as processing at the highest level since their averages determined their DOP scores. Next, percentages of participants who processed at each respective level of DOP are presented. Subsequently, DOP is investigated at length via the construct of awareness. Finally, comments that did not qualify for the DOP schema but that are still of interest to how participants process are presented.

Research question 1a: *ser* versus *estar*. Table 28 displays examples of participants’ TAs on the copulas by level of DOP regarding the copulas. Translations are provided in square brackets where necessary. Some of the metalinguistic WCF participants mentioned the metalinguistic abbreviations, the full list of which can be consulted in Appendix J. For quicker reference, the meanings of the specifically mentioned abbreviations are provided in footnotes. Capitalized words or letters within words represent participants’ emphasis.
Table 28

Examples of Participants’ Think Aloud Protocols on Ser versus Estar per Depth of Processing Level

<table>
<thead>
<tr>
<th>Level</th>
<th>Ser versus estar</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High DOP</strong></td>
<td>“Estaba again. That I was with. I guess if you’re with someone it’s estar and not ser? Estaba.” (Use: location, direct)</td>
</tr>
<tr>
<td></td>
<td>“Cita romántica [Romantic date] uhuh oh oh because it’s talking about my state of being so estaaaaba. Romántica estaba muy tímida [Romantic I was very shy]. Good.” (Use: description, metalinguistic)</td>
</tr>
<tr>
<td></td>
<td>“La entrevista estuvió en un café. Estuvió? [The interview was in a café. Was?] Um. Estuvió I feel like that’s not changing so- oh but it is location so maybe that is correct. So estuvió en un café.” (Use: location, control)</td>
</tr>
<tr>
<td><strong>Medium DOP</strong></td>
<td>“Estaban presentes. E- not eran presen-.” (Use: location, direct)</td>
</tr>
<tr>
<td></td>
<td>“V.E.¹ is equivalent verb so I need ESTÁ.” (Use: location [as evidenced by particular sentence she was correcting], metalinguistic)</td>
</tr>
<tr>
<td></td>
<td>“Maybe it’s just fue.” (Use: location, indirect)</td>
</tr>
<tr>
<td><strong>Low DOP</strong></td>
<td>“Fffffffuee.” (Use: location, direct)</td>
</tr>
<tr>
<td></td>
<td>“Soy. I don’t know what’s wrong with soy nervoso² [I am nervous].” (Use: description, metalinguistic)</td>
</tr>
<tr>
<td></td>
<td>“Jeff es.” (Use: description, indirect)</td>
</tr>
</tbody>
</table>

Application of the DOP scores reflected that participants in the direct and metalinguistic WCF groups processed at low, medium, and high levels of DOP for ser and estar and the

---

¹ V.E. stands for incorrect verb.
² Nervoso is not a typographical error. The participant pronounced nervioso as nervoso in this instance.
indirect WCF participants processed at a low level of DOP. One control group participant
verbalized his thoughts regarding *ser* versus *estar*. He processed at a high level of DOP for this
dichotomy only. Table 29 shows the percentage of participants, by group, who processed at each
level of DOP for *ser* versus *estar*.

Table 29

*Depth of Processing Results for Ser versus Estar per Written Corrective Feedback Type*

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Direct WCF</th>
<th>n</th>
<th>Metalinguistic WCF</th>
<th>n</th>
<th>Indirect WCF</th>
<th>n</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>High DOP</td>
<td>2</td>
<td>22.22%</td>
<td>2</td>
<td>22.22%</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>100.00%</td>
</tr>
<tr>
<td>Medium DOP</td>
<td>2</td>
<td>22.22%</td>
<td>3</td>
<td>44.44%</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Low DOP</td>
<td>7</td>
<td>63.63%</td>
<td>4</td>
<td>33.33%</td>
<td>5</td>
<td>100.00%</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Note.* The dashes demonstrate that indirect participants’ DOP scores did not qualify at high or
medium levels and that no control participants thought aloud regarding *ser* versus *estar*.

Of the participants who verbalized their thoughts on *ser* versus *estar*, the majority (63.63%) of
the direct WCF participants and all of the indirect WCF participants processed at a low level.
The majority (44.44%) of the metalinguistic WCF participants processed at a medium level and
the one control participant who thought aloud regarding this dichotomy processed at a high level.

One way of probing further into DOP is level of awareness. In the (I)SLA field, it has
been argued that higher levels of awareness can lead to better performance in subsequent tests
(e.g., Leow, 1997, 2001a, 2012; Hsieh, Moreno, & Leow, 2015). Awareness at the level of
understanding is the highest level of awareness wherein learners arrive at the underlying rule of
the target. While awareness is utilized to probe more deeply into DOP, the results must be
viewed with caution due to such low $n$-sizes. Five participants (direct WCF, $n = 2$; metalinguistic WCF, $n = 2$; control, $n = 1$) earned a high DOP score for *ser* versus *estar*.

In one particular direct WCF participant’s case, he reached awareness at the level of understanding in the context of description: “That’s a feeling$^3$ so I’m gonna use *estar* and not *ser*. *Estaba entusiasmado* [I was excited],” and location of people/objects: “*Estaba* again. That I was with. I guess if you’re with someone it’s *estar* and not *ser? Estaba,*” but not for location of events: “*Mi primera cita romántica. Fue*. So ag- (exhales) I don’t know. I guess in the past for locations I’d use *ser.*” Here, he hypothesized about the location of a romantic date, but he did not arrive at the correct rule. The other direct WCF participant who qualified at a high DOP reached awareness at the level of understanding for *estar* temporary change: “Instead of *fui* it should be *estuve*. I was very sad. Yeah it wasn’t- yeah. That’s not permanent.” He did not TA about the other uses. One high-DOP metalinguistic WCF participant reached awareness at the level of understanding for *estar* temporary change: “What is V.E. again? V.E. is the wrong verb there. Okay so she WAS a certain state so it’s feelings that would be *estar* that would not be *ser*. So it’s *estaba.*” In the other instances, neither high-DOP metalinguistic participant verbalized a rule or thought aloud regarding certain uses of *ser* and *estar* at all. The high-DOP control participant reached awareness at the level of understanding for *ser* permanent characteristics: “And ya know what I’m gonna do? In this case she’s serious but she’s not always serious so I’m gonna say ESTÁ,” but not for location: “*La entrevista estuvo en un café. Estuvo? Um. Estuvo* I feel like that’s not changing so- oh but it is location so maybe that is correct. So *estuvo en un café.*” In this example, we can also see his use of the rule from *estar* temporary change to approach solving this problem.

$^3$ Although this participant is linking the use of *estar* with feelings and emotions, this is technically the correct rule since feelings and emotions can be interpreted as a temporary change.
Table 30 shows the number of high DOP-scoring participants who reached awareness at the level of understanding, or not, for the respective uses of *ser* versus *estar*. It also reflects where these participants either did not verbalize a rule, but were still classified as high DOP based on the schema and scoring system, or did not TA regarding certain uses.

Table 30

*High Depth of Processing and Awareness at the Level of +/- Understanding for Ser versus Estar*

<table>
<thead>
<tr>
<th></th>
<th>Direct WCF</th>
<th>Metalinguistic WCF</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>ser</em> permanent description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Understanding</td>
<td>-</td>
<td>-</td>
<td>1 of 1</td>
</tr>
<tr>
<td>–Understanding</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No rule provided</td>
<td>-</td>
<td>1 of 2</td>
<td>-</td>
</tr>
<tr>
<td>No TA</td>
<td>2 of 2</td>
<td>1 of 2</td>
<td>-</td>
</tr>
<tr>
<td><em>estar</em> temporary change</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Understanding</td>
<td>2 of 2</td>
<td>1 of 2</td>
<td>1 of 1</td>
</tr>
<tr>
<td>–Understanding</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No rule provided</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No TA</td>
<td>-</td>
<td>1 of 2</td>
<td>-</td>
</tr>
<tr>
<td><em>ser</em> location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Understanding</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>–Understanding</td>
<td>1 of 2</td>
<td>-</td>
<td>1 of 1</td>
</tr>
<tr>
<td>No rule provided</td>
<td>-</td>
<td>1 of 2</td>
<td>-</td>
</tr>
<tr>
<td>No TA</td>
<td>1 of 2</td>
<td>1 of 2</td>
<td>-</td>
</tr>
<tr>
<td><em>estar</em> location</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Understanding</td>
<td>1 of 2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>–Understanding</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No rule provided</td>
<td>-</td>
<td>1 of 2</td>
<td>-</td>
</tr>
<tr>
<td>No TA</td>
<td>1 of 2</td>
<td>1 of 2</td>
<td>1 of 1</td>
</tr>
</tbody>
</table>

*Note.* Dashes represent instances where the data did not qualify for a given category.

With regard to *ser* permanent characteristics, the control participant was the only participant to provide a rule and his score remained at 100.00% for this use across the three
drafts. As for *estar* temporary change, the four participants who reached awareness at the level of understanding all improved their scores from Draft 1 to Draft 2. Descriptively, this improvement did not necessarily last over time, as two participants’ scores returned to their Draft 1 scores and two participants’ scores decreased but still remained above their Draft 1 scores. The only participant to provide a rule for *ser* location was the control participant and he did not reach awareness at the level of understanding. His score, nonetheless, stayed the same from Draft 1 to Draft 2 and descriptively improved from Draft 2 to Draft 3. Lastly, for *estar* location the participant who reached awareness at the level of understanding improved his score from Draft 1 to Draft 2 but returned to his original Draft 1 score at the time of Draft 3.

One case of interest is that of a direct WCF participant whose overall DOP score was medium for *ser* versus *estar*, but who reached awareness at the level of understanding in her one instance of high DOP. (Her other instances of DOP were at low and medium levels so her average ended up at a medium). She reached awareness at the level of understanding for *estar* temporary change: “ESTABA desinhibida. Okay because we’re using *estar* here because it’s not permanent characteristic. It’s a STATE of being. That makes sense. Complicated.” Like two of the four participants who processed at a high level of DOP +awareness at the level of understanding, this medium DOP +awareness at the level of understanding participant improved her score from Draft 1 to Draft 2 but returned to her original score by the time of Draft 3.

**Research question 1b: preterit versus imperfect.** Table 31 displays examples of participants’ TAs by level of DOP regarding past tense aspect. Translations are provided in square brackets where necessary. Some of the metalinguistic WCF participants mentioned the metalinguistic abbreviations, the full list of which can be consulted in Appendix J. For quicker
reference, the meanings of the specifically mentioned abbreviations are provided in footnotes.

Capitalized words or letters within words represent participants’ emphasis.

Table 31

*Examples of Participants’ Think Aloud Protocols on the Preterit versus Imperfect per Depth of Processing Level*

<table>
<thead>
<tr>
<th>Level</th>
<th>Preterit versus Imperfect</th>
</tr>
</thead>
<tbody>
<tr>
<td>High DOP</td>
<td>“So another um imperfect so just changing all the verbs to imperfect which makes sense because it’s completed past act- NOT complete. It’s uh ongoing past action.” (Direct)</td>
</tr>
<tr>
<td></td>
<td>“So V.A. indicates that I used the wrong verb. But I didn’t use the wrong verb. Um. Let’s just try um. Fui? [I went?] Oh no it’s not the wrong verb I’m sorry. It’s telling me it shouldn’t be imperfect it should be preterit. So FUE al concierto. [S/he went to the concert].” (Metalinguistic)</td>
</tr>
<tr>
<td>Medium DOP</td>
<td>“Vivía [lived/used to live] not preterit. That’s imperfect.” (Direct)</td>
</tr>
<tr>
<td></td>
<td>“Tuve [I had] is wrong. Tenía [I had].” (Metalinguistic)</td>
</tr>
<tr>
<td></td>
<td>“Hizo [s/he did/made] is right. También hizo [s/he also did/made].” (Indirect)</td>
</tr>
<tr>
<td>Low DOP</td>
<td>“Compramosssss.” (Direct)</td>
</tr>
<tr>
<td></td>
<td>“Cuando teneemmmmnnia una entrevista. [When I had an interview].” (Metalinguistic)</td>
</tr>
<tr>
<td></td>
<td>“Y no charló. Uh charlaba. [And s/he didn’t chat. S/he would/used to chat].” (Indirect)</td>
</tr>
</tbody>
</table>

The direct and metalinguistic WCF groups processed at low, medium, and high levels of DOP and the indirect WCF participants processed at low and medium levels. Table 32 shows the

---

4 V.A. stands for verb aspect.
percentage of participants, by group, who processed at each level of DOP for the preterit versus imperfect. The control group does not appear because none of these participants verbalized their thoughts regarding this dichotomy.

Table 32

_Depth of Processing Results for the Preterit versus Imperfect per Written Corrective Feedback_

<table>
<thead>
<tr>
<th>Type</th>
<th>n</th>
<th>Direct WCF</th>
<th>n</th>
<th>Metalinguistic WCF</th>
<th>n</th>
<th>Indirect WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>High DOP</td>
<td>2</td>
<td>22.22%</td>
<td>4</td>
<td>66.67%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Medium DOP</td>
<td>3</td>
<td>33.33%</td>
<td>1</td>
<td>16.67%</td>
<td>1</td>
<td>20.00%</td>
</tr>
<tr>
<td>Low DOP</td>
<td>4</td>
<td>44.44%</td>
<td>1</td>
<td>16.67%</td>
<td>4</td>
<td>80.00%</td>
</tr>
</tbody>
</table>

_Note._ The dash demonstrates that indirect participants’ DOP scores did not qualify at a high level.

The majority (44.44% and 80.00%, respectively) of the direct and indirect WCF participants processed the preterit versus imperfect at a low level. Most (66.67%) of the metalinguistic WCF participants processed this dichotomy at a high level and none of the control participants verbalized their thoughts regarding the preterit versus imperfect.

Six participants (direct WCF, n = 2; metalinguistic WCF, n = 4) earned a high DOP score for the preterit versus imperfect dichotomy. Unlike the copulas, the prompt was not designed to elicit specific uses of the preterit and imperfect so these past tense aspects were not broken down further. On the other hand, the past tense was not as open a choice as the copulas were at times,
i.e. participants produced action verbs instead of copulas when answering questions posed with “to be,” so the embedded nature of the preterit versus imperfect dichotomy elicited comparative TA instances in the two cases in which rules were verbalized. One participant mentioned: “And then this one is asistía so another um imperfect so just changing all the verbs to imperfect which makes sense because it’s completed past act- NOT complete. It’s uh ongoing past action,” and the other said, “Me vestía I guess that’s over time so it’s not preterit it’s imperfect. Me ves- -tía.” The two participants to arrive at awareness at the level of understanding were in the direct WCF group and they verbalized the correct rule, comparing the preterit and the imperfect when changing their productions from the erroneous preterit to the imperfect. None of the high-DOP metalinguistic WCF participants verbalized any rules, correct or incorrect. Table 33 shows the number of high DOP-scoring participants who reached awareness at the level of understanding for the preterit versus the imperfect.

**Table 33**

*High Depth of Processing and Awareness at the Level of +/- Understanding for the Preterit versus Imperfect*

<table>
<thead>
<tr>
<th></th>
<th>Direct WCF</th>
<th>Metalinguistic WCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preterit versus imperfect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+Understanding</td>
<td>2 of 2</td>
<td>-</td>
</tr>
<tr>
<td>–Understanding</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>No rule provided</td>
<td>-</td>
<td>4 of 4</td>
</tr>
<tr>
<td>No TA</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

One of the participants who reached awareness at the level of understanding by arriving at the correct rule improved his score from Draft 1 to Draft 2 but scored below his original score on
Draft 3 because he overproduced the imperfect. The other participant to reach awareness at the level of understanding improved his score from Draft 1 to Draft 2 and returned to his original score by the time of Draft 3. Of the high DOP participants who did not verbalize any rules regarding past tense aspect, two of them improved from Draft 1 to Draft 2 and while their scores decreased from Draft 2 to Draft 3, they were descriptively higher on Draft 3 than in Draft 1. One of the participants’ scores improved from Draft 1 to Draft 2 and the score was maintained through Draft 3. Lastly, the fourth high DOP participants’ score decreased from Draft 1 to Draft 2 and was returned to the original score by the time of Draft 3.

Participants also made comments about the respective WCF they received, the errors they committed on a more general level or metacognition to monitor their progress, and strategies they employed while making changes in the second draft. Table 34 shows examples of these comments. Translations are provided in square brackets where necessary.
Table 34

Examples of Participants’ Comments or Actions per Written Corrective Feedback Type

<table>
<thead>
<tr>
<th>Comments about WCF</th>
<th>Comments about Errors and Metacognition</th>
<th>Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td><strong>“Ser versus estar. Feel like that’s my biggest mistake.”</strong></td>
<td>“All right. So I’m starting to get tired so I’m gonna switch and go to the shorter paragraph that I have at the end and fix that one first then I’ll go back.”</td>
</tr>
<tr>
<td></td>
<td><strong>“Again I said <em>era</em> instead of <em>estaba</em>. So that’s obviously a problem area.”</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>“I mean I’m changing the- I’m doing a lot of uh conjugations. They were all like I noticed that I was using like <em>eres</em> and it should have been <em>estar</em> and I don’t really like know when to use the imperfect when I’m writing. My writing like- when I- when I hear someone speak, I hear the emphasis like I hear <em>está</em> I can relate it to like a situation like I think about like it’s an ongoing action or like what’s completed you know what I mean? I don’t have to like think about it but when I’m writing, I don’t know how to determine if it’s imperfect or preterit and...”</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>“Did I forget that there’s even the verb <em>estar</em>?”</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>“So seems like a lot of imperfect instead of uh past tense. Preterit.”</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>“I think it’s just a lot of the imperfect and past word things and then articles.”</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>“I guess I’m just making the same mistakes in different places? Don’t know.”</strong></td>
<td></td>
</tr>
<tr>
<td>Comments about WCF</td>
<td>Comments about Errors and Metacognition</td>
<td>Strategies</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>&quot;Mi amiga sin embargo FUE. [My friend however WENT] Oh god. I think we should put it on autocorrect for imperfect. I am very confused about when to use it.&quot;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Um V.E. What’s the V.E. again? Um. <em>Verbo equivocado.</em> Okay so ERAN um I’ve had this a bunch of times. See I actually like this notation. I had a note earlier where the notation was confusing but actually it’s very helpful when it comes to (someone in lab coughs, can’t hear this word) ’cause then I can just refer back to the other notes that I’ve had. Um so like for example if I get a V.E. and I’ve seen the V.E. before I can scroll up a few sentences and then see the fix that I made within context so it’s extremely helpful in saving a lot of time when I make the same mistake over and over again.&quot;</td>
<td>&quot;Oh it has to do with verbs. Every time V.A. came up I just did that wrong. V.A. (Looks at symbol sheet). Oh like the wrong. I feel stupid. Okay. Yeah there isn’t much of a difference between V.E. and V.A. Yo (pause) <em>iba?</em> [I used to go?]&quot;</td>
<td>(Types &quot;V.A.&quot; or &quot;V.E.&quot; as placeholders throughout her draft when she doesn’t know what the code means/how to address it) “Ah this is stressing me out.”</td>
</tr>
<tr>
<td>&quot;V.A. I’m guessing that’s not imperfect that’s preterit again. Let’s look at the notes. Yep. Again. I used the wrong tense. My most common thing is the V.A.’s here. Um I’m not using the preterit / imperfect tenses in the right situations.”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;It has to do with verbs. Every time V.A. came up I just did that wrong. V.A. (Looks at symbol sheet). Oh like the wrong. I feel stupid. Okay. Yeah there isn’t much of a difference between V.E. and V.A. Yo (pause) <em>iba?</em> [I used to go?]&quot;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Why are these underlined?”</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>“The first thing I see is a lotta lines. Yay.”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>-</td>
<td>&quot;I should also, next time (mumbles) conjugate preterit verbs. I’m not sure if I did all of them correctly. I think I did most of them correctly um.”</td>
</tr>
<tr>
<td>&quot;I’m gonna wordreference up a bunch of <em>palabras de transición.</em>&quot;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. Dashes represent categories where participants did not comment.*
Table 35 shows the percentage of participants, by WCF type, that made the certain types of comments.

Table 35

*Commentary per Written Corrective Feedback Type*

<table>
<thead>
<tr>
<th></th>
<th>Direct WCF</th>
<th>Metalinguistic WCF</th>
<th>Indirect WCF</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>About WCF</td>
<td>-</td>
<td>20.00%</td>
<td>100.00%</td>
<td>-</td>
</tr>
<tr>
<td>About Errors, Metacognition</td>
<td>87.50%</td>
<td>40.00%</td>
<td>-</td>
<td>33.33%</td>
</tr>
<tr>
<td>Strategies</td>
<td>12.50%</td>
<td>40.00%</td>
<td>-</td>
<td>66.67%</td>
</tr>
</tbody>
</table>

*Note.* Dashes represent categories where participants did not comment.

The majority (87.5%) of the direct WCF participants who made comments during Time 2 of the study made comments regarding their errors and monitored their progress via metacognition. The other 12.50% made comments about their strategies to work through Draft 2. None of them commented about the WCF itself. Some (20.00%) of the metalinguistic WCF participants commented about the codes and symbols, and the rest (40.00% and 40.00%, respectively) commented about their errors and monitored their progress as well as commenting on their strategies while revising. All of the indirect WCF participants who made comments did so regarding the underlines. As for the control participants, like the direct WCF group, none of them commented about the WCF. Dissimilarly, the majority (66.67%) of the participants to comment did so on their strategies and the rest (33.33%) commented about their errors and monitored their progress while working through Draft 2.
**Research question one summary.** From a qualitative standpoint regarding how participants processed WCF, direct and metalinguistic WCF participants processed at high, medium, and low levels of DOP when revising their errors in the production of *ser* versus *estar* and the preterit versus imperfect. Indirect WCF participants processed at low and medium levels of DOP when addressing *ser* versus *estar* and the preterit versus imperfect. One control participant processed at a high level of DOP when revising errors in the production of *ser* versus *estar* and none of the control participants verbalized thoughts regarding errors in the production of the preterit versus imperfect.

When DOP was investigated more profoundly via the construct of awareness, sporadic patterns were observed. They cannot be interpreted with confidence, as the only a few participants qualified for high DOP and even less qualified for awareness at the level of understanding.

Comments about the WCF, errors, and overall strategies also emerged from the TA protocols. Among these three categories, direct WCF participants commented exclusively on their errors whereas metalinguistic WCF participants commented on all three aspects. Indirect WCF participants commented on the WCF itself and control participants commented on their overall strategy to address the content feedback.

**Research question two.** Research question two (RQ2 asked whether there was a relationship between level of depth of processing of written corrective feedback and adult L2 Beginning Spanish learners’ written production accuracy of *ser* versus *estar* (RQ2a) and the preterit versus imperfect past tense aspects (RQ2b). To address RQ2a and RQ2b, the difference in the scores obtained between Drafts 1 and 3 for both *ser* and *estar* and the preterit versus
imperfect were submitted with the DOP scores to an R-based program, langtest.jp, created by Dr. Atsushi Mizumoto at Kansai University.

**Research question 2a: ser versus estar.** The DOP scores were submitted with the differences in accuracy scores from Draft 1 to Draft 3 per participant were submitted to the correlation test on langtest.jp for *ser* versus *estar*. Table 36 displays the correlation between DOP scores and *ser* versus *estar* score.

Table 36

<table>
<thead>
<tr>
<th></th>
<th>DOP Score and Change from Draft 1 to Draft 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>n</em></td>
</tr>
<tr>
<td></td>
<td>26</td>
</tr>
</tbody>
</table>

In Plonsky and Oswald’s (2014) meta-analysis, they suggest that SLA researchers should interpret correlation coefficients (*r*) in the following sets: small 0.25, medium 0.40, and large 0.60. Considering these classifications, there was a small correlation (*r* = 0.152, *p* = 0.675) between DOP score and accuracy score for the copulas. This result must be interpreted with caution, though, as the *n*-size is quite small and the correlation coefficient is not statistically significant.

**Research question 2b: preterit versus imperfect.** As for the preterit versus imperfect, the DOP scores were submitted to the correlation test on langtest.jp with the differences in accuracy scores from Draft 1 to Draft 3 per participant for the preterit versus imperfect. Table 37 displays the correlations between DOP scores and this dichotomy score.
Table 37  

*Correlations between Depth of Processing Score and Preterit versus Imperfect Score*

<table>
<thead>
<tr>
<th></th>
<th>DOP Score and Change from Draft 1 to Draft 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
</tbody>
</table>

There was a small negative correlation ($r = -0.050$, $p = 0.833$) between DOP score and accuracy score for past tense aspect. This result must be interpreted with caution, though, as the $n$-size is quite small and the correlation coefficient is not statistically significant.

**Research question two summary.** With regard to correlations between level of DOP and accuracy scores of *ser* and *estar* and the preterit versus imperfect, little confidence can be placed on the outcome due to small $n$-sizes. With respect to the copulas, there was a small correlation between DOP score and accuracy. In other words, the higher the level of DOP the higher the accuracy score at the time of Draft 3. In terms of past tense aspect, there was a small negative correlation between DOP score and accuracy. In this case, the higher the level of DOP the lower the accuracy score at the time of Draft 3. This was evident in some cases when participants erroneously overproduced one aspect over the other, despite having processed deeply at the time of Draft 2.

**Research question three.** Research question three asked whether type of unfocused written corrective feedback (direct, indirect, metalinguistic) has a differential effect on adult L2 Beginning Spanish learners’ subsequent written production accuracy of *ser* versus *estar* (RQ3a) and the preterit versus imperfect past tense aspects (RQ3b). If so, does the effect on
accuracy on each respective target dichotomy last over 2 weeks? To address RQ3, scores obtained for ser and estar and imperfect and preterit on Draft 1, Draft 2, and Draft 3 were submitted separately to a 4 x 3 repeated-measures ANOVA in SPSS, version 22, in which Type of WCF (Direct, Metalinguistic, Indirect, and control) was entered as a between-subject factor while Time (Draft 1, Draft 2, and Draft 3) was entered as a within-subject factor. First, descriptive statistics, ANOVA tables, and, when necessary, Scheffé post hoc analyses and effect sizes are presented first for ser versus estar, followed by similar data for imperfect and preterit.

**Research question 3a: ser versus estar.** The instances of each copula ser and estar were too few to analyze separately by use and were therefore collapsed. Table 38 displays the mean scores in percentages for ser versus estar across all three drafts.

Table 38

<table>
<thead>
<tr>
<th>WCF Type</th>
<th>Draft 1</th>
<th>Draft 2</th>
<th>Draft 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Direct</td>
<td>12</td>
<td>46.89 (11.03)</td>
<td>98.99 (2.56)</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>11</td>
<td>47.02 (12.90)</td>
<td>65.62 (22.72)</td>
</tr>
<tr>
<td>Indirect</td>
<td>9</td>
<td>56.74 (9.25)</td>
<td>56.61 (12.06)</td>
</tr>
<tr>
<td>Control</td>
<td>8</td>
<td>43.17 (19.26)</td>
<td>43.82 (19.31)</td>
</tr>
</tbody>
</table>

Table 39 displays a summary of the findings from the between-subject ANOVA for ser versus estar.
Table 39

Summary of Between-Subject Analysis of Variance for Ser versus Estar

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCF</td>
<td>3</td>
<td>2488.116</td>
<td>5.441</td>
<td>0.003</td>
<td>0.312</td>
</tr>
<tr>
<td>Error</td>
<td>36</td>
<td>457.325</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *p* is significant at ≤ 0.05.

The between-subject, repeated measures ANOVA exposed a significant main effect for Type of WCF (*F*(1, 36) = 5.441, *p* = 0.003). In order to probe more deeply into the results between groups, a post hoc Scheffé analysis was run. Table 40 displays the Scheffé results between groups for their scores on *ser* versus *estar* at the time of Draft 2.

Table 40

Post hoc Scheffé for Ser versus Estar Scores on Draft 2

<table>
<thead>
<tr>
<th>WCF Type</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th><em>p</em></th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct vs. Metalinguistic</td>
<td>33.370</td>
<td>6.603</td>
<td>0.000</td>
<td>[14.007, 52.732]</td>
</tr>
<tr>
<td>Direct vs. Indirect</td>
<td>42.374</td>
<td>6.975</td>
<td>0.000</td>
<td>[21.920, 62.828]</td>
</tr>
<tr>
<td>Direct vs. Control</td>
<td>55.163</td>
<td>7.220</td>
<td>0.000</td>
<td>[33.990, 76.335]</td>
</tr>
<tr>
<td>Metalinguistic vs. Indirect</td>
<td>9.004</td>
<td>7.110</td>
<td>0.662</td>
<td>[-11.845, 29.853]</td>
</tr>
<tr>
<td>Metalinguistic vs. Control</td>
<td>21.793</td>
<td>7.350</td>
<td>0.047</td>
<td>[0.239, 43.347]</td>
</tr>
<tr>
<td>Indirect vs. Control</td>
<td>12.789</td>
<td>7.687</td>
<td>0.440</td>
<td>[-9.751, 35.328]</td>
</tr>
</tbody>
</table>

Note. *p* is significant at ≤ 0.05.

The Scheffé analysis revealed that at the time of Draft 2, the direct WCF group
outperformed the other two experimental and control groups ($p = 0.000$) and the metalinguistic WCF group outperformed the control group ($p = 0.047$). No significant differences were found between groups at the time of Draft 3, as shown in Table 41.

Table 41

*Post hoc Scheffé for Ser versus Estar Scores on Draft 3*

<table>
<thead>
<tr>
<th>WCF Type</th>
<th>Mean Difference</th>
<th>Std. Error</th>
<th>$p$</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct vs. Metalinguistic</td>
<td>4.659</td>
<td>5.610</td>
<td>0.875</td>
<td>[-11.792, 21.110]</td>
</tr>
<tr>
<td>Direct vs. Indirect</td>
<td>-5.719</td>
<td>5.927</td>
<td>0.818</td>
<td>[-23.098, 11.660]</td>
</tr>
<tr>
<td>Direct vs. Control</td>
<td>7.881</td>
<td>6.135</td>
<td>0.651</td>
<td>[-10.108, 25.870]</td>
</tr>
<tr>
<td>Metalinguistic vs. Indirect</td>
<td>-10.378</td>
<td>6.041</td>
<td>0.411</td>
<td>[-28.092, 7.336]</td>
</tr>
<tr>
<td>Metalinguistic vs. Control</td>
<td>3.222</td>
<td>6.245</td>
<td>0.966</td>
<td>[-21.535, 21.535]</td>
</tr>
<tr>
<td>Indirect vs. Control</td>
<td>13.600</td>
<td>6.531</td>
<td>0.246</td>
<td>[-5.550, 32.750]</td>
</tr>
</tbody>
</table>

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

Table 42 shows a summary of the findings of the within-subject ANOVA.

Table 42

*Summary of Within-Subject Analysis of Variance for Ser versus Estar*

<table>
<thead>
<tr>
<th></th>
<th>$df$</th>
<th>Mean Square</th>
<th>$F$</th>
<th>$p$</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>39.307</td>
<td>0.865</td>
<td>0.358</td>
<td>0.023</td>
</tr>
<tr>
<td>Time*WCF</td>
<td>3</td>
<td>27.532</td>
<td>0.606</td>
<td>0.615</td>
<td>0.048</td>
</tr>
<tr>
<td>Error</td>
<td>36</td>
<td>45.431</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p is significant at $\leq 0.05$
The within-subject, repeated measures ANOVA revealed that there was no main effect for Time ($F(1, 36) = 0.865, p = 0.358$), and no significant interaction between Time and Type of WCF ($F(1, 36) = 0.606, p = 0.615$). Figure 6 captures the performances of Type of WCF over Time on *ser* and *estar*.

Figure 6

**Ser versus Estar Scores by Written Corrective Feedback Type**

![Graph showing scores by Written Corrective Feedback Type](image)

Despite the nonsignificant result from the ANOVA, Figure 6 merits further investigation via Cohen’s $d$ effect sizes, which reflect the size of the differences between mean scores (Plonsky & Oswald, 2014). The effect sizes are of particular interest to probe into the direct and metalinguistic WCF groups, as the mean scores appeared different between the times of Draft 1 and Draft 2 especially. Plonsky and Oswald (2014) propose a field-specific range for Cohen’s $d$, which is as follows: 0.60 is small, 1.00 is medium, and 1.40 is large. Table 43 presents the within-group effect sizes for *ser versus estar*. 
Table 43

*Within-Subject Effect Sizes for Ser versus Estar*

<table>
<thead>
<tr>
<th></th>
<th>Draft 1 to Draft 2</th>
<th>Draft 2 to Draft 3</th>
<th>Draft 1 to Draft 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct WCF</strong></td>
<td>( t (15.94); p = 0.00 )</td>
<td>( t (14.66); p = 0.00 )</td>
<td>( t (1.05); p = 0.31 )</td>
</tr>
<tr>
<td></td>
<td>( d = 1.05; [-8.63, -4.38] )</td>
<td>( d = 0.91; [4.00, 7.96] )</td>
<td>( d = 0.17; [-4.60, 13.99] )</td>
</tr>
<tr>
<td></td>
<td>MEDIUM</td>
<td>SMALL/MEDIUM</td>
<td>SMALL</td>
</tr>
<tr>
<td><strong>Metalinguistic WCF</strong></td>
<td>( t (2.36); p = 0.04 )</td>
<td>( t (2.41); p = 0.03 )</td>
<td>( t (0.019); p = 0.99 )</td>
</tr>
<tr>
<td></td>
<td>( d = 0.20; [1.89, 35.31] )</td>
<td>( d = 0.21; [0.08, 1.98] )</td>
<td>( d = 0.18; [-10.98, 11.19] )</td>
</tr>
<tr>
<td></td>
<td>SMALL</td>
<td>SMALL</td>
<td>SMALL</td>
</tr>
<tr>
<td><strong>Indirect WCF</strong></td>
<td>( t (0.03); p = 0.98 )</td>
<td>( t (0.14); p = 0.89 )</td>
<td>( t (0.14); p = 0.89 )</td>
</tr>
<tr>
<td></td>
<td>( d = 0.22; [-0.99, 1.01] )</td>
<td>( d = 0.22; [-1.07, 0.93] )</td>
<td>( d = 0.22; [-1.06, 0.93] )</td>
</tr>
<tr>
<td></td>
<td>SMALL</td>
<td>SMALL</td>
<td>SMALL</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>( t (0.07); p = 0.95 )</td>
<td>( t (0.01); p = 0.99 )</td>
<td>( t (0.05); p = 0.96 )</td>
</tr>
<tr>
<td></td>
<td>( d = 0.25; [-1.11, 1.04] )</td>
<td>( d = 0.25; [-1.07, 1.08] )</td>
<td>( d = 0.25; [-1.10, 1.05] )</td>
</tr>
<tr>
<td></td>
<td>SMALL</td>
<td>SMALL</td>
<td>SMALL</td>
</tr>
</tbody>
</table>

\(^*p \) is significant at \( \leq 0.05 \)

There was a medium effect size \( (d = 1.05, p = 0.00) \) from Draft 1 to Draft 2 and a small-to-medium effect size \( (d = 0.91, p = 0.00) \) from Draft 2 to Draft 3 within the direct WCF group.

Since the effect sizes were statistically significant in these two cases, one can have confidence in them. As for the remaining within-group combinations, there were statistically insignificant small effect sizes, suggesting little difference in their mean scores for *ser* versus *estar* over time.

**Research question 3b: preterit versus imperfect.** Descriptive statistics are presented in Table 44 for the preterit versus imperfect scores across all three drafts.
Table 44

*Descriptive Statistics for Preterit versus Imperfect Scores per Written Corrective Feedback Type*

<table>
<thead>
<tr>
<th>WCF Type</th>
<th>Draft 1</th>
<th>Draft 2</th>
<th>Draft 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Direct</td>
<td>11</td>
<td>76.92 (15.12)</td>
<td>99.20 (2.04)</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>9</td>
<td>71.00 (22.43)</td>
<td>89.25 (13.21)</td>
</tr>
<tr>
<td>Indirect</td>
<td>9</td>
<td>73.20 (10.89)</td>
<td>74.18 (12.99)</td>
</tr>
<tr>
<td>Control</td>
<td>5</td>
<td>76.72 (5.36)</td>
<td>76.72 (5.36)</td>
</tr>
</tbody>
</table>

Table 45 displays a summary of the findings from the between-subject ANOVA for the preterit versus imperfect.

Table 45

*Summary of Between-Subject Analysis of Variance for Preterit versus Imperfect*

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>WCF</td>
<td>3</td>
<td>421.697</td>
<td>0.849</td>
<td>0.478</td>
<td>0.078</td>
</tr>
<tr>
<td>Error</td>
<td>30</td>
<td>496.722</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. p is significant at ≤ 0.05.

The between-subject, repeated measures ANOVA did not yield a significant main effect for type of WCF, $F(1, 30) = 421.697, p = 0.478$, between groups. In other words, there were no differential performances between all experimental groups.
Table 46 shows a summary of the findings of the within-subject ANOVA.

Table 46

Summary of Within-Subject Analysis of Variance for Preterit versus Imperfect

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>p</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>1</td>
<td>0.218</td>
<td>0.001</td>
<td>0.974</td>
<td>0.000</td>
</tr>
<tr>
<td>Time*WCF</td>
<td>3</td>
<td>215.705</td>
<td>1.039</td>
<td>0.390</td>
<td>0.094</td>
</tr>
<tr>
<td>Error</td>
<td>30</td>
<td>207.627</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. p is significant at ≤ 0.05.

The within-subject, repeated measures ANOVA revealed no main effect for Time \((F (1, 30) = 0.001, p = 0.974)\) and there was not a significant interaction between Time and Type of WCF \((F (1, 30) = 1.039, p = 0.390)\).

Figure 7 captures the performances of Type of WCF over Time on the preterit versus imperfect.

Figure 7

Preterit versus Imperfect Scores by Written Corrective Feedback Type
Although the ANOVA revealed no significant result, Figure 7 appears to warrant deeper probing via effect sizes for the direct and metalinguistic WCF groups predominantly. Table 47 presents the within-subject effect sizes for the preterit versus imperfect.

Table 47

*Within-Subject Effect Sizes for the Preterit versus Imperfect*

<table>
<thead>
<tr>
<th></th>
<th>Draft 1 to Draft 2</th>
<th>Draft 2 to Draft 3</th>
<th>Draft 1 to Draft 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct WCF</td>
<td>( t (4.84); p = 0.00 ) ( d = 0.28; [-3.17, -0.96] ) SMALL</td>
<td>( t (3.51); p = 0.01 ) ( d = 0.23; [0.49, 2.50] ) SMALL</td>
<td>( t (0.79); p = 0.44 ) ( d = 0.18; [-0.56, 1.23] ) SMALL</td>
</tr>
<tr>
<td>Metalinguistic WCF</td>
<td>( t (2.10); p = 0.06 ) ( d = 0.25; [-2.05, 0.07] ) SMALL</td>
<td>( t (0.99); p = 0.34 ) ( d = 0.23; [-0.54, 1.48] ) SMALL</td>
<td>( t (0.074); p = 0.47 ) ( d = 0.23; [-1.36, 0.66] ) SMALL</td>
</tr>
<tr>
<td>Indirect WCF</td>
<td>( t (0.17); p = 0.86 ) ( d = 0.22; [-1.08, 0.92] ) SMALL</td>
<td>( t (0.46); p = 0.66 ) ( d = 0.22; [-0.79, 1.22] ) SMALL</td>
<td>( t (0.30); p = 0.77 ) ( d = 0.22; [-0.86, 1.14] ) SMALL</td>
</tr>
<tr>
<td>Control</td>
<td>( t (0.00); p = 1.00 ) ( d = 0.40; [-1.46, 1.46] ) SMALL</td>
<td>( t (0.00); p = 1.00 ) ( d = 0.40; [-1.46, 1.46] ) SMALL</td>
<td>( t (0.00); p = 1.00 ) ( d = 0.40; [-1.46, 1.46] ) SMALL</td>
</tr>
</tbody>
</table>

*p is significant at \( < 0.05 \)

There were small effect sizes for all of the within-group combinations, indicating little difference in their mean scores for the preterit versus imperfect score over time. The only effect sizes in which one can place confidence are those that are statistically significant: the direct WCF group’s Draft 1 to Draft 2 and Draft 2 to Draft 3 changes in scores.

*Research question three summary.* The results must be interpreted with caution due to low n-sizes per group. The between-subject repeated measures ANOVA for *ser* versus *estar* yielded a significant main effect for Type of WCF and a post hoc Scheffé analysis revealed
significant differences at the time of Draft 2 between the direct and metalinguistic; direct and indirect; direct and control; and the metalinguistic and control groups. Overall, the direct WCF group outperformed the other three groups and the metalinguistic WCF group outperformed the control group on Draft 2. However, these significant differences did not last over two weeks.

The within-subject, repeated measures ANOVA revealed that there was no main effect for Time and no significant interaction between Time and Type of WCF. In other words, the immediate effects receiving WCF on ser versus estar in the direct and metalinguistic WCF conditions did not remain over a period of two weeks.

Due to the low n-size per cell and the apparent patterns displayed in Figure 6, effect sizes were calculated. There was a medium effect size from Draft 1 to Draft 2 and a small-to-medium effect size from Draft 2 to Draft 3 within the direct WCF group. In these two cases, the effect sizes were statistically significant and it can be argued with confidence that the direct WCF participants’ ser versus estar accuracy score increased from Draft 1 to Draft 2 and decreased from Draft 2 to Draft 3.

As for the preterit versus imperfect, the between-subject repeated measures ANOVA did not reveal a significant main effect for type of WCF, that is, all experimental groups, including the control, performed similarly on both the immediate and delayed posttests, perhaps due to the initial high pretest scores.

The within-subject repeated measures ANOVA did not reveal a significant change over time for the preterit versus imperfect. Although the descriptive statistics show an increase from Draft 1 to Draft 2 in the direct and metalinguistic WCF groups, they decreased from Draft 2 to Draft 3 and returned to roughly the same scores with which they began at the time of Draft 1.
Overall, it appears that WCF did not have a differential effect on this target dichotomy’s scores over time for all groups.

Once again, due to the low n-size per group and the ostensible patterns displayed in Figure 7, effect sizes were calculated. The only significant effect sizes for the mean scores were in the direct WCF group’s scores for the preterit versus imperfect. When looking at the mean scores, it can be argued with some degree of certainty that the direct WCF participants’ preterit versus imperfect accuracy score increased from Draft 1 to Draft 2 and decreased from Draft 2 to Draft 3.
Chapter V: Discussion and Conclusion

This chapter presents the discussion and conclusion of the current study. Reflections are made on the limitations and subsequent approaches for future research pertaining to written corrective feedback in the ISLA field are proposed.

Discussion

**Research question one.** The first research question asked how adult L2 Beginning Spanish learners process unfocused written corrective feedback on *ser* versus *estar* (RQ1a) and the preterit versus imperfect past tense aspects (RQ1b). TA protocols provided online evidence of learners’ DOP and thought processes while working with the respective WCF. The following discussion involves interpretation, but with caution due to low n-sizes.

The direct and metalinguistic WCF participants processed at high (22.22% of both groups), medium (22.22% and 44.44%, respectively), and low (63.63% and 33.33%, respectively) levels of DOP for *ser* versus *estar*. The indirect WCF participants processed at a low (100%) level of DOP. As for the control group, the single control participant who processed *ser* versus *estar* at a high DOP level scored just below the cut-off point of 65% (with a 64.29%) on Draft 1 and while he maintained this dichotomy score on Draft 2, he improved at the time of Draft 3 (with a 78.57%). In this participant’s case, his improvement may be attributed to processing *ser* versus *estar* at a high DOP level and awareness at the level of understanding.

It may be argued that direct WCF does not promote deep processing of specific targets because it is, in a way, a form of spoon-feeding and does not involve the need to process the correction further. Moreover, the design of the current study was such that all participants used the respective WCF while making changes for Draft 2. This can easily explain why these
particular participants’ scores skyrocketed on Draft 2. None of the direct WCF participants made comments about the feedback itself but half of them did reflect on their overall errors while working with the WCF. They either skimmed the WCF and assessed their problem areas before beginning the revision, i.e. “Looks like a lot of past tense. Looks like a lot of my mistakes where I used past tense instead of the imperfect,” or they reflected throughout Draft 2, i.e. “All of these are imperfect versus preterit. Whoops.” This participant’s comment was made roughly halfway through her revision. While the direct WCF participants did not necessarily process the target dichotomies deeply, it was apparent that they were able to zoom out and see the larger picture regarding their errors.

After the study, they were asked to reflect on the feedback they had received through a set of debriefing questions. In response to whether or not they would change anything about the direct WCF, there were several different responses. Some of them mentioned:

Participant 1: “Nothing. I like to know what I did wrong, not have to figure it out.”

Participant 3: “I noticed I was misusing the preterit/imperfect a lot, and it would have been nice to have a short little explanation as to why my usage was incorrect in each case.”

---

5 Additional analyses were run without the Direct WCF group to ascertain whether there were statistically significant differences between and within the metalinguistic, indirect, and control groups. All scores from all three drafts were submitted to 3 x 3 repeated measures ANOVAs. For ser versus estar, the between-subject repeated measures ANOVA revealed no main effect for Type of WCF ($F(1, 25) = 1.967, p = 0.161$) and the within-subject repeated measures ANOVA revealed no significant interaction between Time and Type of WCF ($F(1, 25) = 0.018, p = 0.982$). For the preterit versus imperfect, the between-subject repeated measures ANOVA revealed no main effect for Type of WCF ($F(1, 20) = 0.538, p = 0.592$) and the within-subject repeated measures ANOVA revealed no significant interaction between Time and Type of WCF ($F(1, 20) = 3.201, p = 0.062$). These results differ from the original results that included the Direct WCF group in that no significant difference was found between groups for ser versus estar, which may be due to the Direct WCF group’s near-ceiling performance on Draft 2 that appeared to have caused the significant result.
Participant 4: “This would be more complicated, but maybe if I made the same exact mistake multiple times only correct the first one and make me think and apply it to the following times.”

Participant 6: “Nothing, I thought that it was very clear and concise. It helped me correct my mistakes.”

Participant 11: “For the most part, feedback was about grammar, spelling and verb tense but some advice on writing style and what can be added might be nice.”

Participant 1 seemed to appreciate the spoon-feeding style this WCF offers. Meanwhile, Participant 3’s comment about wanting a brief explanation to accompany the direct corrections, alluded to the metalinguistic explanations utilized in previous WCF research (e.g. Bitchener, 2008; Shintani et al., 2014). Her hint at wanting to better understand her errors was justified by her processing at medium and high levels, respectively, on the copulas and past tense aspect. However, her scores on both targets hit ceiling (100%) on Draft 2 and returned to the original scores on Draft 3, indicating that while she attempted to process deeply, and in the case of the copulas she reached awareness at the level of understanding, she still did not maintain her improved score over time. Participant 4 wanted a little less spoon-feeding, by receiving WCF only on the first occurrence of a given error and not on subsequent errors of the same type. This participant’s ser versus estar score improved to ceiling on Draft 2 and decreased to nearly her original score from Draft 1, but still ended up a few points higher. She had scored above the cut-off point for past tense aspect so her scores were not included in the analysis. Overall, despite Participant 4 wanting to be corrected less frequently, she processed at a low DOP level and did not maintain her 100% Draft 2 copula score over time. Participant 11 seemingly requested to be in the control group in a future WCF study. Overall, the direct WCF group’s scores likely
increased to near-ceiling on average in Draft 2 because they were not only given the specific
corrections but use the corrections during the revision, unlike in previous WCF studies (e.g.,
Shintani et al., 2014; Stefanou & Révész, 2016) wherein participants were given 5 minutes to
read through the feedback and then it was taken away before they began the next assessment
task. The current study differs from these previous studies in that it was more closely related to
pedagogical practices, which allow students to use feedback while they revise their
compositions.

As for metalinguistic WCF, it does seem to promote deeper processing because it
involves problem-solving or the need to resolve the correction. Of these participants, five of
them had not used the symbol sheet before. Of the five metalinguistic WCF participants who
verbalized their thoughts (about either target dichotomy), one expressed that she was having
difficulty with the revision in general: “Ah this is stressing me out.” This participant also typed
the metalinguistic abbreviations, i.e. “V.E.” for incorrect verb or “V.A.” for verb aspect,
throughout her second draft as placeholders so she could proceed and return to them, as
evidenced by the Echo360 video recording. Despite her low DOP score, her preterit versus
imperfect score descriptively improved on Draft 2, but then decreased by Draft 3 to a slightly
lower score than her original score. Her ser versus estar score was not included in the analysis
because her DOP score fell into the low-medium buffer. Nonetheless, her copula score doubled
from Draft 1 to Draft 2 but returned to the original score by Draft 3. Another participant wrestled
with the revision, seemingly overwhelmed, “There are so many things wrong with this I don’t
know what to do. Huuuuuuuuh (sigh). Where am I?” This comment is not necessarily about the
WCF but the participant was clearly struggling at this point in time, 37 minutes into his revision.
He made changes for 11 minutes after making that comment and finished Draft 2 after nearly 48
minutes total. In his case, it is unknown whether he had seen the symbol sheet in his class or not. Regardless of his low DOP score for *ser* versus *estar*, he improved over the course of all three drafts (46% to 54% to 69%). Although he sounded overwhelmed by the revision and processed for the copulas at a low level, he was able to improve and keep improving from Draft 1 to Draft 3.

While the metalinguistic WCF apparently promoted deeper processing among participants in this group, as evidenced by the percentage of participants (22.22% high and 44.44% medium for the copulas and 66.67% high and 16.67% medium for past tense aspect), they did not maintain their improved scores by the time of Draft 3 as a group. Regardless of how they processed the WCF, participants’ scores overall did not improve over time. In response to the debriefing question regarding whether or not they would change anything about the metalinguistic WCF they received, they reflected:

Participant 17: “I don’t like all the different signs to memorize and it doesn’t really help me not make the same mistakes.”

Participant 20: “I always have a hard time knowing what I did wrong. I see what the teacher says is wrong, but I have no clue what to change it to.”

Participant 22: “Having explanations in English as well (for what certain marks indicate).”

Participant 24: “The process of figuring out what I had done incorrectly was time consuming because I had to find the correction in the key provided. I think that making the corrections more direct would have shortened my correction time.”

Participant 26: “More specific, less abbreviations.”
Participant 28: “I like the feedback, I just do better if someone explains it to me in person.”

Participant 29: “I don’t think there’s anything major that needs to be changed. Sometimes it was hard to figure out why something needed correcting, but I don’t think it was an issue with the feedback, I think it was more an issue of me.”

Participants 17, 22, 24, and 29 had not seen the symbol sheet before the time of the study. Participant 17 struggled with her revision in terms of total time of task, which was roughly an hour and a half. Participant 22, while processing at a high DOP level and reaching awareness at the level of understanding for both target dichotomies, mentioned he would have preferred to have the metalinguistic abbreviation explanations in English on the symbol sheet. Although he was the only participant to mention the language on the symbol sheet, he had a point. One speculation is that the metalinguistic WCF participants would have benefited from abbreviation explanations given in their L1. Regardless of delivery language, this participant processed deeply and arrived at the correct rules for the copulas and past tense aspect. Participant 24 felt as though the symbol sheet extended her typical revision time, a hunch that was generally justified by the metalinguistic WCF group’s average time on task, which will be discussed below. Participant 25 spent about a half an hour revising but his mention of wanting more detail was interesting because he qualified for a high DOP, but only because he spent much time processing the target and not because he attempted to verbalize a rule. Participant 29 took full ownership of her work when she reflected on the WCF. She was another participant who had never seen the symbol sheet before. Participants 20, 26, and 28 had used the symbol sheet in their classes three times over the course of the semester prior to the current study. The former two expressed their confusion and wanting more elaboration on their errors while the latter would have enjoyed
participating in previous studies where teacher conferences were held (e.g., Bitchener et al., 2005; Ferris et al., 2013). Overall, the metalinguistic participants mentioned they had some trouble with the symbols, but this group processed at medium and high levels primarily for both target dichotomies.

As for the indirect WCF, this type of feedback may be assumed to promote some depth of processing as it indicates errors have been committed without further guidance. However, these participants appeared not to have processed above a low level of DOP, except for one participant’s single instance of medium DOP for the preterit versus imperfect, because they simply did not know what kinds of changes to make where the underlines appeared. During the revision on Draft 2, only one participant expressed her seemingly overwhelmed feeling via the online measure, i.e. “The first thing I see is a lotta lines. Yay,” and “I still don’t know why some of these are underlined,” (the latter comment was made after 25 minutes of revising). Although an offline measure, the debriefing questions served as a supplement to shed some light onto the participants’ reactions to the WCF. When asked what they would change about the WCF they received in the study, they responded in the following ways:

Participant 34: “More specificity in my mistakes.”

Participant 40: “I love the way my errors were underlined to show what I did wrong throughout my paper but I felt as though if I knew what the error was it would be great. Like if it was due to a wrong conjugation or wrong verb or wrong word, I would like to know which part was wrong.”

Participant 41: “Nothing, I completely understood the feedback given.” (This participant indeed improved slightly from Draft 1 to Draft 2, but returned to her original score by the time of Draft 3).
Participant 42: “It was hard to figure out what I got wrong from the underlines but it was helpful in trying to figure out what I am not good at.”

Participant 43: “I recognized the grammar mistakes I normally make and now I remember those mistakes for sure.” He had already scored above cut-off for *ser* versus *estar* and this score therefore was not analyzed in this case, but his preterit versus imperfect score increased at Draft 2 and then decreased to an even lower score than his original score at Draft 3.

Participant 46: “More specific in terms of what was wrong originally.”

Overall, 50% of the indirect WCF participants said they liked the feedback they received in the study and 50% said they did not. Whether they liked the WCF or not, many of them (62.5%) craved more specific feedback to guide them in their revisions. They all processed at a low level of DOP, except in one participant’s case, who qualified for medium DOP of past tense aspect but who only verbalized once regarding this target. Generally, indirect WCF appeared to promote a low depth of processing and no substantial improvement across the three drafts.

All of the control participants commented on connector words, one made a list of connector words he might want to incorporate, and 33.33% mentioned looking up vocabulary items (although 100% of them used www.wordreference.com in their revision, as evidenced by the Echo360 video recording). In response to the debriefing question regarding whether or not they would change the type of feedback they received in the study, some of the control participants mentioned:

Participant 50: “While helpful, I found myself not knowing where to start. In my current Beginning Spanish class, my professor will give us in depth feedback, i.e. This sentence needs to be changed to this, etc.”
Participant 52: “I would like more details on grammatical changes.”

Participant 53: “I would’ve appreciated a little more feedback on weird grammatical tenses and maybe some example suggestions as to how to make the composition flow better - it seemed a little difficult to write out the whole thing without mechanically going through each point one by one.”

Participant 60: “I wouldn’t change anything. I am just beginning Spanish so the suggestion I got of adding more transitions was difficult for me, even though it was a simple request.”

Like the indirect WCF participants and some of the metalinguistic WCF participants, the control participants craved more guidance and specificity from the feedback. However, Participant 60 was an exception when she mentioned she thought the feedback was at the appropriate level for her abilities.

Of interest, too, is the actual number of errors (not the percentages) on *ser* versus *estar* from Draft 1 that participants were dealing with and the frequency with which they verbalized their thoughts regarding these errors during the revision at Time 2. Of the participants who verbalized their thoughts, direct WCF participants spent an average of 19 minutes, 59 seconds revising their drafts, committed an average of 8.36 errors on Draft 1, and they thought aloud an average of 6.64 times regarding *ser* versus *estar*. The metalinguistic WCF participants spent an average of 42 minutes, 54 seconds on Draft 2, which was a bit longer than the other groups. A plausible explanation may be that it is part of the nature of metalinguistic feedback that it will take a bit longer given the need to decipher the feedback via codes and symbols. Within the subset of metalinguistic WCF participants (*n* = 9) who verbalized their thoughts regarding *ser* versus *estar*, two participants had never used the symbol sheet. The average time on task for the
participants who had never used the symbol sheet was 41 minutes, 56 seconds while the average time on task for the other seven participants who had become accustomed to the symbol sheet was 43 minutes, 11 seconds. Combined, these nine participants thought aloud an average of 5 times each about *ser* versus *estar* and had committed an average of 7.78 errors on Draft 1. The indirect WCF participants spent an average of 35 minutes, 12 seconds revising, committed an average of 8.40 errors on Draft 1, and verbalized their thoughts about this dichotomy an average of 4.8 times per participant. The one control participant to TA regarding the copulas spent 22 minutes, 55 seconds revising, committed of 5 errors on Draft 1, and thought aloud 3 times about the copulas. Table 48 provides a summary of average time on task, average number of errors, and average number of TA instances about the copulas.

Table 48

*Average Time on Task, Number of Errors, and Think Aloud Protocols for Ser versus Estar per Written Corrective Feedback Type*

<table>
<thead>
<tr>
<th>Written Corrective Feedback Type</th>
<th>n</th>
<th>Average Time on Task</th>
<th>Average Number of Errors</th>
<th>Average Number of TA Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>11</td>
<td>19:59</td>
<td>8.36</td>
<td>6.64</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>9</td>
<td>42:54</td>
<td>7.78</td>
<td>5.00</td>
</tr>
<tr>
<td>Indirect</td>
<td>5</td>
<td>35:12</td>
<td>8.4</td>
<td>4.80</td>
</tr>
<tr>
<td>Control</td>
<td>1</td>
<td>22:55</td>
<td>5</td>
<td>3.00</td>
</tr>
</tbody>
</table>

*Note.* The average times spent revising differ per group between target dichotomies (also see Table 50) because some participants verbalized about one dichotomy and not the other or vice versa.
Table 48 reflects that participants in the experimental groups committed roughly the same average amount of errors on *ser* versus *estar* and processed for those errors most often in the direct WCF group, followed by the metalinguistic WCF and indirect WCF groups, and finally the control participant who processed his 5 copula errors 3 times. At first glance, the average time on task appears to be different, but a paired-samples t-test revealed (based on the average time on task and participants’ copula scores on Draft 2 and Draft 3) a trend toward significance on Draft 2 and no significant difference between groups on Draft 3. Table 49 presents the t-test results.

Table 49

<table>
<thead>
<tr>
<th>T-test Results for Average Time on Task and Ser versus Estar Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Time on Task &amp; Draft 2 Score</td>
</tr>
<tr>
<td>Time on Task &amp; Draft 3 Score</td>
</tr>
</tbody>
</table>

*Note. p* is significant at ≤ 0.05.

Given the trend towards significance on Draft 2, it appears that time on task as it related to the WCF type may have affected participants’ *ser* versus *estar* scores. Another plausible explanation may be participants’ individual differences, wherein some learners take longer to complete tasks than others.

Although the prompt was designed to elicit more instances of the target dichotomies, participants did not produce *ser* versus *estar* every time they were prompted and when they
produced the copulas, they were not wrong every time since ser is the default and they were indeed prompted to produce ser as often as they were prompted to produce estar.

For the preterit versus imperfect, the direct and metalinguistic WCF participants processed at high (22.22% and 66.67%, respectively), medium (33.33% and 16.67%, respectively), and low (44.44% and 16.67%, respectively) levels of DOP. The indirect WCF participants processed at low and, in one participant’s case, medium levels of DOP. None of the control participants thought aloud regarding the preterit versus imperfect dichotomy. The lack of TAs regarding past tense aspect may be attributed to the high score that many of the control participants started with on this dichotomy and the data therefore were cleansed to exclude the TAs of all participants who had scored at or above the 90% cut-off point on the first draft.

To probe more deeply, one could consider the actual number of errors committed in Draft 1 that participants faced in the Draft 2 revision and the frequency with which they verbalized their thoughts regarding these errors. Among those who thought aloud regarding past tense aspect, the direct WCF participants spent an average of 20 minutes, 23 seconds revising their drafts. They committed an average of 4.70 errors on Draft 1 and thought aloud an average of 6.1 times regarding their preterit versus imperfect errors. Unlike the direct WCF, the metalinguistic WCF participants spent double the amount of time with an average of 52 minutes, 16 seconds on Draft 2, committed an average of 3 errors on Draft 1, and thought aloud an average of 4.8 times about their errors. The indirect WCF participants spent an average of 39 minutes, 7 seconds revising, committed an average of 3 errors on Draft 1, and verbalized their thoughts on the errors they committed on this dichotomy an average of 1 time each. Table 50 provides a summary of average time on task, average number of errors, and average number of TA instances about the past tense aspect dichotomy per WCF type.
Table 50

*Average Time on Task, Number of Errors, and Think Aloud Protocols for the Preterit versus Imperfect per Written Corrective Feedback Type*

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Average Time on Task</th>
<th>Average Number of Errors</th>
<th>Average Number of TA Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>9</td>
<td>20:23</td>
<td>4.70</td>
<td>6.11</td>
</tr>
<tr>
<td>Metalinguistic</td>
<td>6</td>
<td>52:16</td>
<td>3.00</td>
<td>4.83</td>
</tr>
<tr>
<td>Indirect</td>
<td>5</td>
<td>39:07</td>
<td>3.00</td>
<td>1</td>
</tr>
<tr>
<td>Control</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Note.* The average times spent revising differ per group between target dichotomies (also see Table 48) because some participants verbalized about one dichotomy and not the other or vice versa.

Table 50 reflects that participants in the direct WCF group committed the highest average amount of errors on the preterit versus imperfect and processed for those errors most often. The metalinguistic WCF and indirect WCF participants committed the same average number of errors, but the metalinguistic WCF participants processed more often than the indirect WCF participants. It is also of interest to note that the direct WCF and metalinguistic WCF participants’ average number of TA instances is higher than their average number of errors.

Upon further inspection, it was discovered that they processed for the preterit versus imperfect even when revising productions for Draft 2 on which they had not committed an error and hence had not received WCF on Draft 1. Additionally, the average time on task seemed to be different between groups. The metalinguistic WCF participants’ time on task appeared considerably
longer than the other WCF participants’ times. As seen above, Participant 17 had not used the symbol sheet before the study and seemed to struggle with it, as she took about an hour and a half to revise. She verbalized her difficulty with the task during the revision and also expressed her dislike for the symbols in her response to the debriefing questions. Without Participant 17, the metalinguistic WCF group’s average time on task was still descriptively longer than the other groups’ times, at 43 minutes, 2 seconds. With her data included, a paired-samples t-test revealed (based on the average time on task and participants’ aspect scores on Draft 2 and Draft 3) a trend toward significance on Draft 2 and a significant difference between groups on Draft 3. Table 51 presents the t-test results.

Table 51

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>t</th>
<th>p</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time on Task &amp; Draft 2 Score</td>
<td>2</td>
<td>-3.930</td>
<td>0.059</td>
<td>[-110.025, 4.985]</td>
</tr>
<tr>
<td>Time on Task &amp; Draft 3 Score</td>
<td>2</td>
<td>-5.518</td>
<td>0.031</td>
<td>[-62.862, -7.778]</td>
</tr>
</tbody>
</table>

*Note.* The degrees of freedom is 2 because the control group participants did not TA about the preterit versus imperfect and thus were not included in this t-test. *p* is significant at ≤ 0.05.

In light of the significant result of the t-test for Draft 3, it appears that time on task as it related to the WCF type did not affect participants’ preterit versus imperfect scores. This could indicate that more time does not necessarily translate into better performance over time for this dichotomy.
Although the prompt was designed to elicit many instances of the past tense aspect, participants were not wrong every time they produced the preterit and imperfect. Furthermore, many of their mistakes regarding aspect appeared in the copulas and for scoring purposes, these instances were disregarded.

The first research question addressed how learners processed during the revision phase of a composition. These Beginning L2 Spanish learners processed at all three levels of DOP. By group, they processed for *ser* versus *estar* at a high (22.22% of the direct and metalinguistic groups), medium (22.22% of the direct and 44.44% of the metalinguistic), and low (63.63% of the direct, 33.33% of the metalinguistic, and 100% of the indirect participants) levels. The only control participant who processed *ser* versus *estar* did so at a high DOP level. In relation to the preterit versus imperfect, they processed at high (22.22% of the direct and 66.67% of the metalinguistic group), medium (33.33% of the direct, 16.67% of the metalinguistic, and 20% of the indirect group), and low (44.44% of the direct, 16.67% of the metalinguistic, and 80% of the indirect group) levels of DOP. None of the control participants thought aloud regarding the preterit versus imperfect dichotomy.

It could be argued that the metalinguistic WCF in this study promoted deeper processing because it involved deciphering symbols, optionally reading accompanying examples, and figuring out the correction for themselves. The direct WCF was more of a spoon-fed type of feedback but participants processed at overall medium and low levels. The indirect participants almost completely processed at a low level, possibly because they did not know how to approach their revision, as evidenced by their debriefing questions. While the control participants seemed to like the feedback they received, they still mentioned they would have preferred more guidance and grammatical corrections.
Research question two. The second research question asked if there was a relationship between level of depth of processing of written corrective feedback and adult L2 Beginning Spanish learners’ written production accuracy of *ser* versus *estar* (RQ2a) and the preterit versus imperfect past tense aspects (RQ2b).

Given that the *n*-sizes were quite low, not much confidence can be placed in the correlations. There was barely a correlation found between either target dichotomy score and DOP (*r* = 0.152, *p* = 0.675 for the copulas and *r* = -0.050, *p* = 0.833 for past tense aspect). In light of Plonsky and Oswald’s (2014) suggested Pearson’s *r* ranges, these correlations are below the small range of 0.25 and they are not statistically significant. Looking at the raw data, no matter how participants processed, their scores were sporadic from beginning to end. The cut-off points (65% for copulas and 90% for the past tense aspect) to begin with were quite high so the range in scores at the time of Draft 1 included scores from 0.00% to 64.29% for the copulas and from 15.79% to 89.29% for the past tense aspect.

In addition, unlike some previous studies (e.g., Stefanou & Révész, 2015) participants in the current study were not required to re-write their compositions and instead revised their original drafts at Time 3. It would be difficult to control for productions that were comparable to productions in Drafts 1 and 2 without this method. Some of their scores returned to their original Draft 1 scores at the time of Draft 3 possibly because they were provided their original drafts and did not have to produce the targets again. Instead, they had to revise their drafts and decide to make changes or not. That said, it would not have been a fair to compare a participant who made an effort but made the incorrect changes in Draft 3 with a participant who decided not to make changes at all.
**Research question three.** The third research question asked whether type of unfocused written corrective feedback (direct, indirect, metalinguistic) has a differential effect on adult L2 Beginning Spanish learners’ subsequent written production accuracy of *ser* versus *estar* (RQ3a) and the preterit versus imperfect past tense aspects (RQ3b), and if so, does the effect on accuracy on each respective target dichotomy last over 2 weeks (RQ3c)? The following discussion needs to be taken with caution due to low *n*-sizes.

Type of unfocused WCF had a differential effect on the accurate production of *ser* versus *estar*. Specifically, the difference in performance appeared between the direct WCF group and all three of the other groups (metalinguistic, indirect, and control). The direct WCF group outperformed these groups on the copula score at the time of Draft 2. One plausible explanation for the direct WCF group’s success on Draft 2 is that participants essentially had a cheat sheet while making their revisions. The metalinguistic WCF group outperformed the control group on Draft 2 as well. A likely explanation is that they were given hints as to what the errors were while the control group was not provided with any indication of errors.

In terms of previous research, the finding that the unfocused direct WCF group outperformed the other unfocused WCF groups on the revision corroborates part of Chandler’s (2003) Study 2 in that the unfocused direct group outperformed the unfocused indirect + metalinguistic and metalinguistic-only groups, but not completely. Chandler’s unfocused indirect-only group also outperformed these groups. This link needs to be taken lightly because Chandler measured accuracy by the mean number of errors per 100 words rather than targeting specific items or structures. The current study reflected opposing results to some previous studies, as well. For example, Evans et al. (2011), Hartshorn et al. (2010), and Lalande (1982) found that the unfocused metalinguistic group outperformed the unfocused direct group in their
compositions, but they, like Chandler, measured accuracy holistically. They did not include an indirect WCF group or a control group so further comparisons cannot be made. One previous study that the current study’s findings refute in terms of the immediate posttest is Truscott and Hsu (2008) wherein the unfocused indirect WCF group outperformed the control group on the immediate posttest. They measured grammatical and orthographic errors on more of a holistic scale like the studies discussed above.

As for the preterit versus imperfect, type of unfocused WCF did not have a differential effect on participants’ accurate production of this target dichotomy. One possible explanation for the lack of differences in the past tense aspect score is that the cut-off point was set at near-ceiling (90%) because the participants scored quite high from the beginning, leaving little room for improvement overall. Despite collecting data the same week they were formally exposed to the imperfect in their classes, 25 of 61 of them scored above a 90% on Draft 1. As discussed earlier, one reason the participants performed so well on Draft 1 is that they were not true beginners. Although enrolled in Beginning and Introductory I Spanish courses, the language background questionnaire revealed that they had an average of 2.21 years of previous Spanish study. Also, 18 of them had studied Latin or French before entering college for an average of 4.94 years.

In relation to previous research, the finding that there was not a differential effect for type of WCF in this case supports, in part, Robb et al. (1986) but refutes Fathman and Whalley (1990) and Ferris and Roberts (2001). Robb et al. found no difference among groups with accuracy measured by error-free T-units, but they did find that the mean scores improved over time. Regarding previous studies’ findings that were refuted in the current study, Fathman and Whalley found that the unfocused indirect WCF groups outperformed the control group and a
group that received content comments, which would later become the control condition in some WCF studies, (e.g., Ellis et al., 2008; Sheen et al., 2009). Ferris and Roberts found that both the unfocused metalinguistic and unfocused indirect groups outperformed the control group. Both of these latter studies scored accuracy on a holistic scale.

Considering their performances on both dichotomies in the current study, participants potentially had more practice with *ser* versus *estar* because they committed an average of 7.39 errors per participant on Draft 1. Hence, they received more feedback on this target, whereas they had arguably less practice and thus less feedback with the preterit versus imperfect because they committed an average of 3.57 errors per participant on Draft 1, most likely due to their prior knowledge of these target items.

Unfocused WCF did not have a lasting effect (cf. Ashwell, 2000; Ferris, 2006; Robb et al., 1986; e.g., Ferris et al., 2013; Kepner, 1991; Mubarak, 2013; Polio et al., 1998; Semke, 1984; Truscott & Hsu, 2008) over a period of two weeks on these learners’ written production accuracy of the target dichotomies *ser* versus *estar* or the preterit versus imperfect. It is possible that the unfocused WCF overwhelmed the participants insofar as they had a considerable amount of corrections to confront in their revisions (i.e. Ellis et al., 2008). Perhaps, cognitively speaking, the need to divide their effort to address many different errors (and not just errors on the target dichotomies) could have impacted their depth of processing to achieve awareness at the level of understanding of the various uses of the target items. For example, of the five participants who earned high DOP scores for *ser* versus *estar*, four of them reached awareness at the level of understanding, but not necessarily for each use of *ser* and each use of *estar*. One of them stated the correct rule for one use and hypothesized incorrectly over another use. Lastly, one high-processing participant simply did not state a rule and therefore did not qualify for the awareness
analysis. This does not imply that this participant did not understand the uses, but rather she spent time processing and therefore earned the high DOP score and she did not state a hypothesis or rule in her TA. As for the preterit versus imperfect, two of the six high DOP participants stated the correct rule for this target dichotomy and the other four high DOP participants did not hypothesis or state a rule, but rather they spent time processing and therefore qualified for this level of DOP. Even after processing highly and in some cases, arriving at the correct rule, with so many types of feedback to address in unfocused WCF, the ability to retain the rules could have been challenging for some of these participants. Since the current study sought to be as pedagogical as possible, the WCF was unfocused as this is the type of feedback provided in most language classes. Likewise, the current study attempted to provide some explanation for why participants performed the way they did, by provided online process measures. The previous studies discussed above cannot explain what exactly their participants were doing with the feedback.

**Limitations and Future Research**

The current study has several limitations. First, the most imperative change for the future would be to increase the n-size. One rather large limitation stemming from the low n-size is the issue of reactivity. It was not addressed in the current study due to an initial participant pool of 61, which would be inevitably reduced after cleansing. Therefore, in the future it would be ideal to start with more participants and add a group that does not TA. Second, it is unknown whether the act of thinking aloud helped the participants, hindered their performance, or neither. Previous studies on written production (Sachs & Suh, 2007; Suh, 2010; Yanguas & Lado, 2012) also do not shed much light on any potential impact. According to Suh’s (2010) dissertation, thinking aloud did not cause reactivity among her participants. However, Sachs and Polio (2007) reported
negative reactivity during the reformulation phase of their experimental treatment but only in one of two experiments. However, their participants thought aloud in their L2 English. Lastly, Yanguas and Lado (2012) found positive reactivity among their participants. However, while thinking aloud helped them outperform the non-TA groups in terms of error-free units in their writing, no reactivity was found in other facets of their results, namely, number of words, number of words per T-unit, lexical variety, and fluency and lexical complexity. In sum, given the design difference between the current study and the previous ones in light of the focus of corrective (revision vs. production), the addition of a non-TA group is clearly warranted.

Third, the researcher waited to collect data until the participants had studied the targets in their Spanish classes. Otherwise, it would not have been realistic to expect them to produce them in their writing (i.e. Ellis et al., 2008). However, upon grading the pretest, it was discovered that the participants had quite a grasp on both target dichotomies and the cut-off points had to be established at an above-chance level. The preterit versus imperfect dichotomy cut-off point was especially high, leaving little room for substantial improvement to take place. The researcher consulted the language background questionnaire to search for an explanation as to why some participants scored so high on Draft 1. As seen above, it was discovered that the average time these participants had studied Spanish was 2.21 years. Therefore, they were not true beginners. Additionally, 18 participants had studied Latin or French during elementary, middle, and high school for an average of 4.94 years.

Fourth, with regard to the prompt, certain questions conflated copulas and aspect. For example, a question like Where was your first interview? demands that a participant select a copula and a past tense aspect. In future WCF studies that employ composition tasks, the prompts should elicit one linguistic target per question.
Fifth, during data collection, the researcher noticed some of the metalinguistic WCF participants commenting about how they liked the symbols key. Once the sessions concluded, she asked them if they had used this key in their classes throughout the semester. Six of the sixteen participants in this experimental condition had never seen it before, despite its otherwise widespread use in the cross-sectional Spanish program at the university. The metalinguistic WCF group took the longest of all four groups to complete their revisions. This may be attributed to having to locate the symbols and codes on the double-sided key and then reading the corresponding meaning and examples in Spanish.

In future WCF studies, it would be of interest to compare various proficiency levels, individual differences, affect and motivation, task complexity, linguistic targets in more L2 and heritage languages, and to have participants not only revise but also compose new texts longitudinally. With respect to proficiency levels and individual differences, it could be the case that certain learners process particular types of WCF or certain types of linguistic items more effectively than others. With respect to linguistic targets, many have yet to be investigated within the WCF realm, especially in languages other than ESL. Since foreign language learning is not a linear process, it would be of interest to observe, for example, adult L2 Spanish learners’ or heritage learners’ learning of a specific linguistic target and how they process the WCF through a longitudinal study.

Conclusion

In conclusion, this dissertation contributed to the WCF sub-strand of ISLA in that it offers qualitative insight into how adult L2 Beginning Spanish learners process while they interact with WCF. Learners who received unfocused direct or metalinguistic WCF processed at high, medium, and low levels and learners who received indirect WCF processed primarily at a
low level. They commented on their mistakes regarding the target dichotomies, as well, in the
direct, metalinguistic, and control groups. The metalinguistic and indirect participants
commented about the WCF they received, and participants from all groups except the indirect
group commented on the strategies they were employing to confront the revisions.

However, the results yielded barely any correlation between the different levels of
processing, which may be attributed to the nature of unfocused WCF. Despite some participants’
reaching a high level of DOP and, in some cases, arriving at the correct underlying rule for each
of the target dichotomies, little-to-no correlation was found. It is possible that these learners had
too much to correct in their drafts and therefore, regardless of some of them arriving at the
correct rule, were overwhelmed by having to confront many different categories of corrections.
Consequently, the correlations reflected that accuracy, regardless of DOP level, barely changed
from Draft 1 and Draft 3. As a reminder, the scores entered in the analysis were overall accuracy
for each copula and not individual use scores for each copula.

The present study also revealed that participants exposed to unfocused direct WCF
performed significantly better than the other experimental conditions (metalinguistic, indirect,
and control) on the copula score at the time of Draft 2. The metalinguistic WCF group also
outperformed the control group on Draft 2 as well. However, these superior performances were
not maintained two weeks later. With respect to the preterit versus imperfect, type of WCF did
not appear to play a significant role in subsequent performances on both the immediate and
delayed posttests.

Recently, WCF researchers in the ISLA field have fine-tuned their research questions and
WCF provision by focusing on a few specific linguistic targets (e.g., Karimi, 2016; Sheen et al.,
2009; Stefanou & Révész, 2015). Unfocused WCF provision has subsided, with a few recent
exceptions (e.g. Evans et al., 2010; Ferris et al., 2013; Hartshorn et al., 2010) since the 1980s. This dissertation was based on fusing the best of both worlds: to bring a pedagogical emphasis together with a fine-tuned research design.

With regard to writing in language courses, not just composition courses, Polio and Lee (in press) offer, “The evidence that students focus on language as they write...is overwhelming, even if the long-term effects of this focus have not been well documented. This focus suggests that writing-to-learn-language activities should be used in most language classes” (p. 310). Although significant changes were not observed over the few weeks of the experiment, L2 learning is not linear and it takes time.

In conclusion, the qualitative data gathered in this study has shed some light onto the internal processes employed by learners while working with feedback and, hopefully, it is a data collection tool that will be used in future WCF studies. In the words of a participant who responded in the “additional comments” section of the debriefing question set, “I really enjoyed the process of talking out my own writing, as I am finding it help me in my Spanish writing process as I was forced to think about how I think.” Perhaps a better understanding of how our L2 writers process the feedback we provide on their compositions may lead to a better feedback approach that maximizes the value of the feedback in the promotion of L2 development.
APPENDIX A: Language Background Questionnaire

Participant #______

IRB #2015-1288 (pilot study)
IRB #2016-0758 (main study)

LANGUAGE BACKGROUND QUESTIONNAIRE

1. Sex: __________
2. Age: __________
3. Country of birth: __________
4. What is/are your native language(s)? ______________________________________________
5. What language(s) do you speak at home? __________________________________________
   a. If more than one, with whom do you speak each of the languages?
      __________________________________________
6. In what language(s) did you receive the majority of your precollege education?
   __________________________________________
   a. If more than one, please give the approximate number of years for each language.
      Language: __________ Years: __________
      Language: __________ Years: __________
7. Have you studied any foreign languages other than Spanish? Yes_____ No_____ 
   a. If yes, please specify all information including 1) which language, 2) how long, and
      3) in what context(s)
      I have studied ________________ for __________ year(s) Context: ________________
8. Have you ever been to a Spanish-speaking region for the purpose of studying Spanish? Yes_____ No_____ 
   a. If yes, when? ______________________________________________________________________
      Where? ______________________________________________________________________________
      For how long? _______________________________________________________________________
9. Other than the experience mentioned in Question 7, have you ever lived in a situation where you were exposed to a language other than your native language(s) (e.g., by living in a multilingual community; visiting a community for purposes of study abroad or work; exposure through family members, etc.)? Yes_____ No_____
If yes, please provide details below.

<table>
<thead>
<tr>
<th>Experience 1</th>
<th>Experience 2</th>
<th>Experience 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country/region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>From when to when</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. In the boxes below, rate your language ability in each of the languages that you know. Use the following ratings: 0 = poor, 1 = good, 2 = very good, 3 = native/nativelike

<table>
<thead>
<tr>
<th>Language</th>
<th>Listening</th>
<th>Speaking</th>
<th>Reading</th>
<th>Writing</th>
<th>Number of years of study</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11. Have you studied Spanish in school in the past at any of the levels listed below?
   If yes, for how long?
   a. Elementary school:
      Yes____ No____ Less than 1 year____ 1-2 years____ More than 2 years____
   b. Junior high/middle school:
      Yes____ No____ Less than 1 year____ 1-2 years____ More than 2 years____
   c. Senior high school:
      Yes____ No____ Less than 1 year____ 1-2 years____ More than 2 years____
   d. University:
      Yes____ No____ Less than 1 year____ 1-2 years____ More than 2 years____
   e. Other (please specify: ____________):
      Yes____ No____ Less than 1 year____ 1-2 years____ More than 2 years____

12. What year are you in school? Freshman____ Sophomore____ Junior____ Senior____
    Graduate student____ Other____

13. What is your major? ________________________________
APPENDIX B: TA Practice Activity

Think-Aloud Practice

Think aloud in English while you complete the following task of calculating the total cost of your items at Hoya Snaxa while you wait in line to pay. In other words, verbalize aloud or say whatever comes to your mind while you perform this task. You don’t need to explain what you are saying; just talk aloud as you do the task.

You are going to buy:

- 2 packs of gum at 50¢ each =
- Laundry detergent at $5.00 =
- 4 bottles of water at $1.50 each =
- Solo cups at $5.00 =

Total = __________

You have $20. How much will you have left? __________
APPENDIX C: Pilot Study Prompt

COMPOSITION 4

Imagine you’ve been asked by a blogger from a Spanish-language website to post an entry about life as a teenager in your hometown. Now that you’re in college, reflect back on those years. If any of the scenarios below do not apply to your or your friends’ real life experiences, use your imagination!

**In Spanish**, address the following questions about (1) yourself and (2) a friend of the opposite sex:

<table>
<thead>
<tr>
<th>(1) Think of yourself as a teenager:</th>
<th>(2) Think of a friend of the opposite sex during that time:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• How were you normally versus when you were at parties?</td>
<td>• How was your friend normally versus when s/he was at parties?</td>
</tr>
</tbody>
</table>

Now, describe the most fun party you went to during your teenage years (where was it, who was there, what happened, why it was so much fun, etc.)

<table>
<thead>
<tr>
<th>• How were you as a student? How were you at work after school?</th>
<th>• How was your friend as a student? How was s/he at work after school?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• What and where was your favorite thing to do during your free time?</td>
<td>• What and where was your friend’s favorite thing to do during her/his free time?</td>
</tr>
</tbody>
</table>

Now, describe the most embarrassing moment of your teenage years (where was it, who was there, what happened, why it was so embarrassing, etc.)

¡OJO! Remember to talk about yourself and a friend of the opposite sex in the above situations! You may wish to end with a comment about whether or not you and your friend enjoyed your teenage years overall.

Type 200-250 words (approximately ½ to ¾ page, double-spaced, Times 12 font). **You may use wordreference.com ONLY to look up vocabulary words. DO NOT use the conjugation feature or any other websites.**
APPENDIX D: Pilot Study Prompt with Uses of Linguistic Targets

<table>
<thead>
<tr>
<th>(1) Think of yourself as a teenager:</th>
<th>(2) Think of a friend of the opposite sex during that time:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• How were you normally [IMP ser permanent + AGR] versus when you were at parties [IMP estar change + AGR]?</td>
<td>• How was your friend normally [IMP ser permanent + AGR] versus when s/he was at parties [IMP estar change + AGR]?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Now, describe the most fun party you went to during your teenage years (what happened, who was there, why it was so much fun, etc.)</th>
<th>[Open to multiple uses of ser, estar, AGR, PRET, IMP]</th>
</tr>
</thead>
<tbody>
<tr>
<td>• How were you as a student [IMP ser permanent + AGR]? How were you at work after school [IMP estar change + AGR]?</td>
<td>• How was your friend as a student [IMP ser permanent + AGR]? How was s/he at work after school [IMP estar change + AGR]?</td>
</tr>
<tr>
<td>• What [IMP ser permanent] and where [IMP ser location] was your favorite thing to do during your free time?</td>
<td>• What [IMP ser permanent] and where [IMP ser location] was your friend’s favorite thing to do during her/his free time?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Now, describe the most embarrassing moment of your teenage years (what happened, who was there, why it was so embarrassing, etc.)</th>
<th>[Open to multiple uses of ser, estar, AGR, PRET, IMP]</th>
</tr>
</thead>
</table>

Note. AGR stands for “agreement,” in this case gender agreement, PRET stands for “preterit,” and IMP stands for “imperfect”.
APPENDIX E: Accent Mark Typing Instructions

Keystrokes for Spanish accent marks on a Mac already IN SPANISH MODE:

á, é, í, ó, ú [ then the letter (a, e, i, o, u)
ü OPTION + u, then
ñ ; ñ
í SHIFT + <
i SHIFT + 1 (number 1)
APPENDIX F: Email Script to Instructors

I am writing to remind you that I will be meeting with your Beginning Spanish class this week in the Language Laboratory (ICC 227). Your students will complete a composition in Spanish that I will analyze for the pilot study component of my dissertation. The goal of the study is to investigate the learning process in a second/foreign language. Please attend all sessions, which will take place during your normal class time. It is important that you attend, as your students will have 30 minutes to complete the writing tasks and then you will conduct class as usual for the remainder of the class period. As for the follow-up sessions, the only difference will be that your students will have less time to complete the revision tasks than they did to complete the compositions. Again, you will conduct class as usual for the remainder of the class period. Another reason to attend is in the event a student decides to opt out of the composition, we will provide her/him with an alternate pedagogical activity to complete while her/his classmates are completing the experimental task and you will assist to monitor her/him.

Secondly, you will not see the two compositions the students write during the data collection period. Instead, I will grade them according to the Department of Spanish & Portuguese’s composition grading guidelines and rubric and report the grades to you. I will send the students’ graded compositions to them. Their grades will not be of any relevance to the study and therefore will not be reported.

I will make it clear to your students that they are not required to participate, but will need to report to class as usual and will complete an alternate activity for a grade to maintain fairness among all students. While their classmates complete the composition, they will complete a pedagogical activity. Once again, you will not be responsible for monitoring your students who complete the experimental activity, but rather only monitor those students who may choose to
opt out of the study. You will provide a grade on this activity in lieu of the composition grade I would have given them had they participated in the study.

I will also provide them with my contact information (amc358@georgetown.edu) should they have any questions or want to discuss their compositions with me once the study has concluded. You may also contact me should you have any questions. Thank you in advance.

Below is the information your students will receive upon entering the laboratory on the first day of data collection:

You are invited to participate in a research study that is being conducted by a graduate student in the Department of Spanish and Portuguese, Georgetown University. The goal of the study is to investigate the learning process in a second/foreign language.

Participation in this study is entirely voluntary at all times. You can choose not to participate at all or to leave the study at any time. Regardless of your decision, there will be no effect on your relationship with the researcher or any other consequences.

You are being asked to take part in this study because you are an English speaker with minimum linguistics background. If you agree to participate, you will be recorded throughout the experiment session as you complete a task on the computer. If you would not like to be recorded, please inform the researcher and you will be excused from participating. The sessions will be approximately 30 minutes or less. You will read a prompt in English and then write and revise a composition in Spanish.

Your responses recorded for this study will remain anonymous and cannot be linked to you in any way. No identifying information about you will be collected at any point during the study, and your data will be identified only with a random ID number.

Data for this study will be kept in password-protected folders on a computer that
is kept under lock and key. While only the researcher (Allison Caras) will have access to the data, she may consult with her professors for discussion.

There are no risks associated with this study. Information collected in this study may benefit others in the future by helping researchers and teachers to better understand language processing and hopefully improve language teaching.

If you have any questions regarding this research project, please contact the investigator, Allison Caras, at amc358@georgetown.edu. If you have any questions about your rights as a research participant, please contact the Georgetown University IRB at (202) 687-6553 or irboard@georgetown.edu.

By participating in the recorded communication tasks, you are indicating your consent to participate in this study. Thank you.
APPENDIX G: Email Recruitment Script to Participants

You are invited to volunteer to participate in a study on cognitive processes. The study is being carried out by a graduate student from the Department of Spanish and Portuguese. The study investigates the process of learning Spanish as a Foreign Language at the first level. Your participation in this study is totally voluntary, and it is not a requirement of your current class. Should you decide to participate in the study, extra credit will be granted to one of your two quizzes and you will receive a gift card to The Corp upon completion of the study. In order to participate in the study, you will need to meet with the researcher for approximately 2.5 hours over three sessions. You will complete a composition in Spanish in the Language Lab. Once again, your participation is voluntary and if you would like to opt out of participating, then you may do so at any time. Please contact Allison Caras at amc358@georgetown.edu if you are interested in participating. Thank you
Imagine you’ve been asked by a blogger from the site ¡Cuéntame algo! to post an entry about life as a teenager in your hometown. Now that you’re in college, reflect back on those years. **Respond to EVERY question.** Check the boxes to mark your progress. If any of the scenarios below do not apply to your or your friend’s real life experiences, use your imagination!

You have roughly 1 hour to write your entry. Type 200-250 words (approximately ½ to ¾ page, double-spaced, Times 12 font). **You may use wordreference.com ONLY to look up vocabulary words. DO NOT use the conjugation feature or any other websites.**

In **Spanish**, address **ALL** of the following questions about (1) yourself and (2) a friend of the opposite sex.

<table>
<thead>
<tr>
<th>(1) Think of yourself as a teenager:</th>
<th>(2) Think of a friend of the opposite sex during that time:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ How were you on a typical day versus on the day of an interview?</td>
<td>□ How was your friend on a typical day versus on the day of an interview?</td>
</tr>
<tr>
<td>□ What was your first interview for?</td>
<td>□ What was her/his first interview for?</td>
</tr>
<tr>
<td>□ Where was the school, company...?</td>
<td>□ Where was the school, company...?</td>
</tr>
<tr>
<td>□ Where was the interview, i.e. in an office, a restaurant, etc.?</td>
<td>□ Where was the interview, i.e. in an office, a restaurant, etc.?</td>
</tr>
<tr>
<td>□ How did it go?</td>
<td>□ How did it go?</td>
</tr>
<tr>
<td>□ How were you normally versus when you were on a first date?</td>
<td>□ How was your friend normally versus when s/he was on a first date?</td>
</tr>
<tr>
<td>□ Where was your first date?</td>
<td>□ Where was her/his first date?</td>
</tr>
<tr>
<td>□ What did you do?</td>
<td>□ What did s/he do?</td>
</tr>
<tr>
<td>□ How were you on an ordinary day versus when you went to your first concert, sports game, other?</td>
<td>□ How was your friend on an ordinary day versus when s/he went to her/his first concert, sports game, other?</td>
</tr>
<tr>
<td>□ Where was the concert/game/other?</td>
<td>□ Where was the concert/game/other?</td>
</tr>
<tr>
<td>□ Who was there?</td>
<td>□ Who was there?</td>
</tr>
<tr>
<td>□ How were you as a student versus when you were at parties?</td>
<td>□ How was your friend as a student versus when s/he was at parties?</td>
</tr>
<tr>
<td>□ Where was the first party you ever attended as a teenager?</td>
<td>□ Where was the first party your friend ever attended as a teenager?</td>
</tr>
<tr>
<td>□ Who was there?</td>
<td>□ Who was there?</td>
</tr>
<tr>
<td>□ Now, describe the most embarrassing moment of your teenage years:</td>
<td></td>
</tr>
<tr>
<td>□ Where were you?</td>
<td></td>
</tr>
<tr>
<td>□ Who was there?</td>
<td></td>
</tr>
<tr>
<td>□ What happened?!</td>
<td></td>
</tr>
</tbody>
</table>

¡OJO! Remember to talk about yourself **and** a friend of the opposite sex in the above situations! Make sure to answer **EVERY** question and feel free to expand on any details.
APPENDIX I: Main Study Prompt with Uses of Targets

<table>
<thead>
<tr>
<th>(1) Think of yourself as a teenager:</th>
<th>(2) Think of a friend of the opposite sex during that time:</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ How were you on a typical day [IMP ser permanent + AGR] versus on the day of an interview [IMP/PRET estar change + AGR]?</td>
<td>□ How was your friend on a typical day [IMP ser permanent + AGR] versus on the day of an interview [IMP/PRET estar change + AGR]?</td>
</tr>
<tr>
<td>□ What was your first interview for? [IMP/PRET]</td>
<td>□ What was her/his first interview for? [IMP/PRET]</td>
</tr>
<tr>
<td>□ Where was the school, company... [IMP/PRET estar location]?</td>
<td>□ Where was the school, company... [IMP/PRET estar location]?</td>
</tr>
<tr>
<td>□ Where was the interview, i.e. in an office, a restaurant, etc. [IMP/PRET ser location]?</td>
<td>□ Where was the interview, i.e. in an office, a restaurant, etc. [IMP/PRET ser location]?</td>
</tr>
<tr>
<td>□ How did it go? [IMP/PRET]</td>
<td>□ How did it go? [IMP/PRET]</td>
</tr>
<tr>
<td>□ How were you as a student [IMP ser permanent + AGR] versus when s/he went to her/his first concert, sports game, other [IMP/PRET estar change + AGR]?</td>
<td>□ How was your friend normally [IMP ser permanent + AGR] versus when s/he was on a first date [IMP/PRET estar change + AGR]?</td>
</tr>
<tr>
<td>□ Where was your first date [IMP/PRET ser location]?</td>
<td>□ Where was her/his first date [IMP/PRET ser location]?</td>
</tr>
<tr>
<td>□ What did you do? [IMP/PRET]</td>
<td>□ What did s/he do? [IMP/PRET]</td>
</tr>
<tr>
<td>□ How were you on an ordinary day versus [IMP ser permanent + AGR] versus when you went to your first concert, sports game, other [IMP/PRET estar change + AGR]?</td>
<td>□ How was your friend on an ordinary day versus [IMP ser permanent + AGR] versus when s/he went to her/his first concert, sports game, other [IMP/PRET estar change + AGR]?</td>
</tr>
<tr>
<td>□ Where was the concert/game/other [IMP/PRET ser location]?</td>
<td>□ Where was the concert/game/other [IMP/PRET ser location]?</td>
</tr>
<tr>
<td>□ Who was there [IMP/PRET estar location]?</td>
<td>□ Who was there [IMP/PRET estar location]?</td>
</tr>
<tr>
<td>□ How were you as a student [IMP ser permanent + AGR] versus when s/he was at parties [IMP/PRET estar change + AGR]?</td>
<td>□ How was your friend as a student [IMP ser permanent + AGR] versus when s/he was at parties [IMP/PRET estar change + AGR]?</td>
</tr>
<tr>
<td>□ Where was the first party you ever attended as a teenager [IMP/PRET ser location]?</td>
<td>□ Where was the first party your friend ever attended as a teenager [IMP/PRET ser location]?</td>
</tr>
<tr>
<td>□ Who was there [IMP/PRET estar location]?</td>
<td>□ Who was there [IMP/PRET estar location]?</td>
</tr>
<tr>
<td>□ Now, describe the most embarrassing moment of your teenage years:</td>
<td></td>
</tr>
<tr>
<td>□ Where were you? [IMP/PRET estar location]</td>
<td></td>
</tr>
<tr>
<td>□ Who was there? [IMP/PRET estar location]</td>
<td></td>
</tr>
<tr>
<td>□ What happened?! [IMP/PRET]</td>
<td></td>
</tr>
</tbody>
</table>

Note. AGR stands for “agreement,” in this case gender agreement, PRET stands for “preterit,” and IMP stands for “imperfect”. It is also clear that in many cases, depending on how each participant responds, the use of the imperfect or the preterit could arise and the researcher will score each composition accordingly.
### APPENDIX J: Metalinguistic Abbreviations and Symbols Key

**ABREVIATURAS Y SÍMBOLOS DE CORRECCIÓN**

<table>
<thead>
<tr>
<th>Key</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>a personal</td>
<td>Todavía no he conocido <em>tu</em> novia → Todavía no he conocido <em>a</em> tu novia</td>
</tr>
</tbody>
</table>
| AC  | acento     | inglés → inglés *(con acento)*  
|     |            | farmacia → farmacia *(sin acento)*  
|     |            | teléfono → teléfono *(con el acento en otro lugar)* |
| ADJ | adjetivo   | Ella es *bien* estudiante → Ella es *buena* estudiante  
|     |            | Es *importante* |
| ADV | adverbio   | Ella habla español muy *bueno*  
|     |            | Jorge aprende fácil → ... *fácilmente* |
| ART | artículo   | Me gusta *biología* → Me gusta *la* biología  
|     |            | *Tecnología cambia diariamente* → *La* tecnología...  
|     |            | Me duele *mi* cabeza → me duele *la* cabeza  
|     |            | Recibí *una* buena educación → Recibí *una* buena educación *(artículo indefinido)* |
| AUX | auxiliary  | Luis se *está* enfermado tres veces este año → Luis se *ha* enfermado...  
|     |            | Los países *mejoraron* mucho → Los países *han* mejorado mucho |
| C   | usar un conector | Estudio mucho. Me gusta salír → Estudio mucho; *sin embargo*, me gusta salir. → *Aunque* estudio mucho, me gusta salir. |
| CONJ | conjunción | antes de *te* vayas → antes de *que* te vayas |
| CONS | construcción | Mi hermano y yo nos gustaba... → A *mi* hermano y a *mí* nos gustaba... |
| ESP | español    | Pennsylvania → Pensilvania |
| G   | género      | *la* tema → *el* tema  
|     |            | *El* es una persona *bueno* → *Él* es una persona *buena* |
| I   | infinitivo | Después de comiendo, voy al museo → Después de *comer*...  
|     |            | Corriendo es bueno para la salud → *Correr* es bueno para la salud  
|     |            | Yo quiero que yo saque una A → Yo quiero *sacar* una A |
| M   | mayúscula  | los estados *unidos* → los Estados Unidos  
| m   | minúscula  | quiero aprender Español → quiero aprender español |
| N   | número      | *mi* hermanos → *mis* hermanos  
|     |            | *mejor* amigos → *mejores* amigos |
| NEG | negativo    | no conozco a alguien → no conozco a nadie |
| ¿OJO! | ¡OJO! (cuidado) | ¿Qué tiempo es? → ¿Qué *hora* es?  
|     |            | He tenido mucho suceso en mi trabajo → He tenido mucho *éxito*...  
|     |            | tener diversión → divertirse  
|     |            | *(Este símbolo puede acompañar OP o ve si el contraste entre palabras/verbos es uno que ya se ha estudiado)* |
| OK  | ignorar     | *Ignora* mi comentario/corrección; *lo* que escribiste está bien |
| ORT | ortografía  | profesora → *profesora*  
|     |            | _abilidad* → habilidad  
|     |            | *anos* → *años* |
| OP  | otra(s) palabra(s) | *significante* → *significativo*  
|     |            | por una mano → por *un* lado; por *una* parte |
| POS | posesivo    | ¿Ud. trajo *tu* carro a la fiesta? → ¿Ud. trajo *su* carro...?  
|     |            | nos clase → *nuestra* clase |
| ¿PSN? | posición | ¿Qué hablaron de? → ¿De qué hablaron? |
La vida nos da oportunidades a la vida.

PRE: preposición

Comí en el restaurante → Comí al restaurante

Estoy aprendiendo a bailar → Estoy aprendiendo ^ bailar

Asiste a una clase → Asiste ^ una clase

PRO: pronombre

Anita las vi a ellas → Anita las vi

Yo le di un regalo a mi mamá → Yo le di un regalo a mi mamá

Les llamé ayer → Los llamé ayer

(pero, hay algunos dialectos que son leístas y usan la primera opción en vez de la segunda)

PUN: puntuación

^Quieres beber algo? → ¿Quieres beber algo?

^Qué bonito! → ¡Qué bonito!

REL: relativo

Hablé con el señor que conozco → Hablé con el señor quien conozco.

u: unir oraciones

El avión es la mejor invención. Es obvio. → Es obvio que el avión es...

Uso el internet todos los días. El internet ha cambiado mi vida. → El internet, lo cual uso todos los días, ha cambiado mi vida.

SUS: sustantivo

Abrí la puerta con mucha fuerza → Abrí la puerta con mucha fuerza

La inteligente de uno es muy importante → La inteligencia de uno...

¿TÍT? : título

(Hay que incluir un título para todas las composiciones)

¿V? : falta el verbo

Yo ^ Almitra. → Yo soy Almitra.

Va: aspecto equivocado

cuando tuve 13 años → cuando tenia 13 años

(El pretérito y el imperfecto difieren en el aspecto)

Vc: conjugación equivocada (o falta de conjugación)

Yo estudié francés en la escuela secundaria → Yo estudié francés...

Puedes hablar conmigo mientras lavas la ropa → mientras lavas la ropa

Ve: verbo equivocado

Almuerzo a las 8 de la mañana → Desayuno a las 8 de la mañana

Encontré a muchas personas en mi viaje → Conocí a muchas personas...

Era mucha gente en la discoteca → Había mucha gente...

Soy cansado → Estoy cansado

Sabes a Ana → Conoces a Ana

vf: forma equivocada

Recientemente, he hablando mucho → Recientemente, he hablado mucho

Vm: modo equivocado

Creo que ella habló demasiado → Creo que ella habló demasiado

Llámame antes de que sales → Llámame antes de que sales

(El indicativo y el subjuntivo difieren en el modo)

Vt: tiempo equivocado

El próximo año fui a Miami → El próximo año iré a Miami

Cuando era niña, siempre jugaba con mi abuelo → ...siempre jugaba...

[]: aplica a

(El comentario/la corrección aplica a lo que está entre corchetes)

---

eliminar

/ /

^: insertar

invertir las palabras

miroja carpeta → mi carpeta roja

¿?: no se entiende

\: nuevo párrafo

Unir palabras: de el → del

* : ver comentario
APPENDIX K: Debriefing Questions

DEBRIEFING QUESTIONS

1. When you have a composition due for Spanish class, do you usually write the whole thing in one session? Yes_____ No_____  
   a. If yes, about how long does it take you from start to finish? _____________________
   b. If no, how many sessions do you usually take? _______________________________
       i. About how long is each session? _________________________________________

2. When you schedule your Spanish composition writing time(s), what do you schedule it/them between, e.g., bio lab, write, go to the gym? ________________________________
       ____________________________________________
       ____________________________________________
       ____________________________________________

3. Do you use online resources when you write your compositions for Spanish class?
   Yes_____ No_____  
   a. If yes, which websites and/or apps do you use, i.e. Google Translate, SpanishDict?
       ____________________________________________
       ____________________________________________
   b. If no, what do you use, i.e. hardcopy dictionary? ____________________________
       ____________________________________________

4. Did you like the feedback you received in the experiment? Yes_____ No_____  
   a. What would you change about the feedback?
       ____________________________________________
       ____________________________________________
       ____________________________________________
5. Additional comments?
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
References


Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the


VanPatten, B. (2010). Some verbs are more perfect than others: Why learners have difficulty and what it means for instruction. *Hispania, 93*(1), 29-38.


Yanguas, I. & Lado, B. (2012). Is thinking aloud reactive when writing in the heritage

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