

ABSTRACT

Title of Dissertation: EXPLICIT WRITTEN CORRECTIVE
FEEDBACK AND LANGUAGE APTITUDE
IN SLA: IMPLICATIONS FOR
IMPROVEMENT OF LINGUISTIC
ACCURACY

Susan Dianne Benson, Doctor of Philosophy,
2016

Dissertation directed by: Dr. Robert DeKeyser, Second Language
Acquisition

Most second language researchers agree that there is a role for corrective feedback in second language writing classes. However, many unanswered questions remain concerning which linguistic features to target and the type and amount of feedback to offer. This study examined two new pieces of writing by 151 learners of English as a Second Language (ESL), in order to investigate the effect of direct and metalinguistic written feedback on errors with the simple past tense, the present perfect tense, dropped pronouns, and pronominal duplication. This inquiry also considered the extent to which learner differences in language-analytic ability (LAA), as measured by the LLAMA F, mediated the effects of these two types of explicit written corrective feedback. Learners in the feedback groups were provided with corrective feedback on two essays, after which learners in all three groups completed two additional writing tasks to determine whether or not the provision of corrective feedback led to greater gains in accuracy compared to no feedback. Both treatment

groups, direct and metalinguistic, performed better than the comparison group on new pieces of writing immediately following the treatment sessions, yet direct feedback was more durable than metalinguistic feedback for one structure, the simple past tense. Participants with greater LAA proved more likely to achieve gains in the direct feedback group than in the metalinguistic group, whereas learners with lower LAA benefited more from metalinguistic feedback. Overall, the findings of the present study confirm the results of prior studies that have found a positive role for written corrective feedback in instructed second language acquisition.

EXPLICIT WRITTEN CORRECTIVE FEEDBACK AND LANGUAGE
APTITUDE IN SLA: IMPLICATIONS FOR IMPROVEMENT OF
LINGUISTIC ACCURACY

by

Susan Dianne Benson

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2016

Advisory Committee:
Professor Robert DeKeyser, Chair
Professor Michael Long
Professor Kira Gor
Dr. Scott Jackson
Professor Melinda Martin-Beltrán, Dean's Representative

© Copyright by
Susan Dianne Benson
2016

Dedication

To my parents, Wynne Wells and Marshall Wells, who never got to see me complete my Ph.D., and to Gavan Benson and Wells Benson, who accompanied me throughout the journey.

Acknowledgements

First, I am very grateful to all of my dissertation committee members for their wisdom and guidance: Robert DeKeyser, Michael Long, Kira Gor, Scott Jackson, and Melinda Martin-Beltrán. It was a privilege to work with all of you.

I am especially thankful to Robert DeKeyser for his thoroughness, kindness and support as my dissertation chair, and for suggesting University Park, the perfect neighborhood when we moved to Maryland.

I am also forever grateful to Scott Jackson for his help with statistical questions and for patiently teaching me how to fit and interpret mixed-effects models in R.

A special thanks to Mike Long for his valuable comments, commas, and feedback on this project and previous projects that helped me to improve as a researcher and writer.

Cathy Doughty and the Center for Advanced Study of Language (CASL) also deserve thanks for providing me with employment and invaluable experience during my first two years of doctoral study. I am very grateful for the support and friendship of my amazing CASL colleagues.

I would also like to acknowledge St. Petersburg College (SPC) for granting me two semesters of sabbatical without which my dissertation project would not have been possible. I am truly thankful for all of my awesome colleagues at SPC who welcomed me into their classrooms and offered marvelous support throughout the data collection. In addition, I owe special thanks to the wonderful and enthusiastic students who participated in the study.

Finally, I am very grateful to my family for their support. I could not have completed my doctorate without their help and encouragement. The greatest thanks of all goes to my husband, Gavan, whose love, support, patience, calm, and perspective, always keeps me moving forward, and to my son, Wells, whose love and kindness keeps me smiling.

Table of Contents

Dedication.....	ii
Acknowledgements.....	iii
Table of Contents.....	v
List of Tables.....	vi
List of Figures.....	vii
Chapter 1: Introduction.....	1
Chapter 2: Review of the Literature.....	4
2.1 Written Corrective Feedback.....	4
2.1.1 Focused and unfocused written corrective feedback.....	5
2.1.2 Types of focused written corrective feedback.....	9
2.1.3 Revision or provision?.....	15
2.1.4 Linguistic form.....	17
2.2 Individual Differences.....	19
2.2.1 Proficiency level.....	20
2.2.2 L1.....	21
2.2.3 Aptitude and motivation.....	22
Chapter 3: Research Questions and Hypotheses.....	28
Chapter 4: Methodology.....	36
4.1 Setting and Participants.....	36
4.2 Target Structures.....	38
4.3 Treatment.....	40
4.4 Procedures.....	41
4.5 Writing Tasks.....	44
4.6 Operationalization of Variables.....	45
Chapter 5: Results.....	47
5.1 Descriptive Statistics for LLAMA F.....	47
5.2 Descriptive Statistics and Analyses for Form-focused Grammar Test.....	48
5.3 Simple Past Tense in the Writing Tasks.....	49
5.3.1 Regular vs. irregular verbs.....	62
5.4 Present Perfect Tense in the Writing Tasks.....	64
5.5 Dropped Pronouns in the Writing Tasks.....	70
5.6 Pronominal Duplication in the Writing Tasks.....	78
5.7 Summary of Findings from Mixed-effects Models.....	86
5.8 Qualitative Data from the Exit Survey.....	87
Chapter 6: Discussion and Conclusions.....	91
6.1 The Effects of Feedback on Accuracy in Subsequent Writing Tasks.....	91
6.2 Error Type and Treatability.....	95
6.3 Interactions Between Feedback Type and Structure.....	99
6.4 Interactions Between Feedback Type and Language-Analytic Ability.....	105
6.5 Conclusions and Directions for Future Research.....	112
Appendices.....	116
Appendix A: Pretest and Posttest Questions.....	116
Appendix B: Writing Prompts.....	122
Appendix C: Exit Questionnaire.....	126
References.....	127

List of Tables

Table 1	<i>Target Structures</i>	29
Table 2	<i>Participants' Native Languages</i>	37
Table 3	<i>Expected Errors Based on L1</i>	39
Table 4	<i>Experimental Procedure</i>	44
Table 5	<i>Descriptive Statistics for LLAMA F</i>	47
Table 6	<i>Descriptive Statistics for Form-focused Grammar Test</i>	48
Table 7	<i>Total Number of Obligatory Contexts for Simple Past at Each Essay Time</i>	51
Table 8	<i>Percentage Correct for Simple Past Tense at Each Essay Time</i>	51
Table 9	<i>Result Summary of Coefficient Estimates, Standard Errors, Associated Wald's z-score, and Significance Level p for the Simple Past Tense Analysis</i>	54
Table 10	<i>Predicted Probabilities of Accurate Use of Simple Past Tense by Group</i>	57
Table 11	<i>Total Number of Trials and Percentage Correct for Regular and Irregular Simple Past</i>	63
Table 12	<i>Total Number of Obligatory Contexts for Present Perfect at Each Essay Time</i>	64
Table 13	<i>Percentage Correct for Present Perfect Tense at Each Essay Time for Each Group</i>	64
Table 14	<i>Result Summary of Coefficient Estimates, Standard Errors, Associated Wald's z-score, and Significance Level p for the Present Perfect Tense Analysis</i>	67
Table 15	<i>Predicted Probabilities of Accurate Use of the Present Perfect Tense</i>	69
Table 16	<i>Descriptive Statistics for Dropped Pronoun Section of Form-focused Grammar Test for Data Subset</i>	71
Table 17	<i>Number of Observations (Obligatory Use of a Subject) at Each Essay Time and Percentage Correct</i>	72
Table 18	<i>Result Summary of Coefficient Estimates, Standard Errors, Associated Wald's z-score, and Significance Level p for the Dropped Pronoun Analysis</i>	75
Table 19	<i>Predicted Probabilities of a Correctly Supplied Pronoun by Group</i>	77
Table 20	<i>Descriptive Statistics for Variables Used in Pronominal Duplication Analysis</i>	79
Table 21	<i>Descriptive Statistics for Pronominal Duplication Section of Form-focused Grammar Test for Data Subset</i>	80
Table 22	<i>Result Summary of Coefficient Estimates, Standard Errors, Associated Wald's z-score, and Significance Level p for Pronominal Duplication Analysis</i>	82
Table 23	<i>Predicted Probabilities for Pronominal Duplication by Group</i>	85
Table 24	<i>Summary of Significant Gains in Accuracy on Immediate and Delayed Written Posttests</i>	86

List of Figures

<i>Figure 1.</i> Correct and incorrect trials with simple past tense for each group at each essay time.....	52
<i>Figure 2.</i> Plot of predicted probabilities of accurate use of simple past tense by group at each essay time.....	57
<i>Figure 3.</i> Effect of LLAMA F at Essay Time 3	60
<i>Figure 4.</i> Correct and incorrect trials with present perfect tense for each group at each essay time.....	65
<i>Figure 5.</i> Plot of predicted probabilities of accurate use of present perfect tense by group at each essay time	70
<i>Figure 6.</i> Estimated marginal means of pretest and posttest scores for pronoun dropping section of form-focused test for subset of learners.....	72
<i>Figure 7.</i> Number of missing pronouns and accurately supplied pronouns for each group at each essay time	73
<i>Figure 8.</i> Plot of predicted probabilities of accuracy for dropped pronouns by group at each essay time.....	78
<i>Figure 9.</i> Estimated marginal means for pretest and posttest scores for pronominal duplication section of form-focused test for subset of learners	81
<i>Figure 10.</i> Plot of predicted probabilities for pronominal duplication by group at each essay time	85
<i>Figure 11.</i> Participants' previous college experience outside U.S.	87
<i>Figure 12.</i> Learner opinion regarding feedback received.....	88
<i>Figure 13.</i> Learner feedback preferences	89
<i>Figure 14.</i> Learner concern about grammatical errors	90

Chapter 1: Introduction

The provision of written corrective feedback by second language writing instructors is a common practice, yet second language (L2) writing researchers are not in agreement concerning the role and effectiveness of corrective feedback (CF). In fact, error correction for L2 writing, and grammar correction in particular, have been hotly debated for several years. Truscott (1996), the most notable opponent of corrective feedback, argued that grammar correction for L2 writing should be “abandoned,” claiming that the research showed it to be both ineffective and even harmful. During the 1980s and early 1990s, the process-oriented perspective from L1 composition became accepted for L2 composition, as many argued that literacy meant students had to discover their voice and ideas and that grammar instruction should not be emphasized. Marking up multiple errors on a student’s paper with red ink would raise the affective filter and stifle the writer’s voice and style. Yet, writing instructors and learners alike still seemed to prefer some form of explicit correction and feedback (Ferris, 1999).

In her review of Truscott’s controversial claims, Ferris (1999) countered that his conclusions were “premature and overly strong” (p. 2) and stressed the need for new research to explore the topic. This article by Ferris started a decade-long debate with Truscott (Ferris, 1999, 2003, 2004, 2006; Truscott, 1996, 1999, 2004, 2007), which has succeeded in sparking a renewed interest in error correction and has generated more controlled research within the field. L2 writing specialists began to counter that a relaxed view of linguistic accuracy or *laissez faire* attitude could be dangerous, as it does not address the linguistic gaps that even advanced learners often have (Leki, 1990; Reid, 1994; Ferris, 1999; Bitchener & Ferris, 2012).

From an acquisition perspective, a flaw found in many early written corrective feedback studies was that they did not consider the effects of feedback on new pieces of writing, but only whether or not feedback helped students to achieve greater accuracy on a second draft (e.g., Ashwell, 2000; Ferris & Roberts, 2001; Sachs & Polio, 2007). In order to determine whether or not written CF will facilitate the acquisition of specific linguistic features, it is necessary to include a posttest, and possibly a delayed posttest, which require new writing within the study design. Recent research has attempted to look at error correction from a perspective of language development and has moved from work concerned mostly with student revisions to more carefully controlled examination of the effect of written CF on learner accuracy for specific linguistic structures. These later investigations have attempted to answer the question of whether or not error correction will ultimately lead to learning and greater accuracy in L2 language development by following research models similar to those used to investigate the various forms of oral error correction. The results of these more recent studies (e.g., Bitchener, 2008; Bitchener & Knoch 2010a, Bitchener & Knoch 2010b; Ellis et al., 2008; Sheen, 2007, 2010; Sheen et al., 2009; Shintani, Ellis, & Suzuki, 2014; Stefanou & Révész, 2015; Van Beuningen et al., 2012) provide consistent evidence that written CF can help learners acquire grammatical features and thus present a serious challenge to Truscott's claim that error correction is ineffective and undesirable. Currently, most L2 researchers agree that there is indeed a role for corrective feedback in L2 writing classes. However, many unanswered questions remain concerning which linguistic features to target and the type and amount of feedback to offer. Learner differences also need to be taken into consideration. Further investigations are needed, therefore, before best practices for providing feedback can be established.

The present study aimed to contribute to this research agenda by examining the differential effects of direct and metalinguistic written feedback on errors with the simple past tense, the present perfect tense, dropped pronouns, and pronominal duplication (e.g., resumptive pronouns from Arabic speakers and topicalization by Spanish speakers). Since several of the written corrective feedback studies conducted since the 1990s have focused on the same linguistic target, English articles, the current study was designed to determine whether or not focused direct and metalinguistic feedback would be effective for treating errors with different structures. One hundred and sixty-five adult ESL learners were matched in sets based on L1 and scores from a test of grammatical inferencing, and then randomly assigned to one of three groups: direct feedback, metalinguistic feedback, or control. Learners in the feedback groups were provided with corrective feedback on two essays, after which learners in all three groups completed two additional writing tasks to determine whether or not the provision of corrective feedback led to greater gains in accuracy compared to no feedback. By including a test of grammatical inferencing, this investigation also considered the extent to which learner differences in language-analytic ability (LAA) mediated the effects of these two types of explicit written corrective feedback.

Chapter 2: Review of the Literature

2.1 Written Corrective Feedback

Many early studies on written corrective feedback focused on the revision process and whether or not feedback was a helpful editing tool for learners (e.g., Ashwell, 2000; Fathman & Whalley, 1990; Ferris 1997), but these studies by design did not seek to determine whether or not feedback would lead to improved accuracy on a new piece of writing.

Other studies, however, focused on the effects of written CF on new pieces of writing to determine whether or not feedback could result in learning. Due to a number of methodological design and execution flaws (e.g., Chandler, 2003; Polio, Fleck, & Leder, 1998), the results from these early studies were mixed (for extensive discussion of these issues see Bitchener & Ferris, 2012; Gu enette, 2007; or Van Beuningen, De Jong, & Kuiken, 2008). For instance, as Truscott (1996) rightly pointed out, studies without control groups cannot offer empirical support for the claim that written CF is effective because the improvements may have nothing to do with the feedback itself. However, most of the studies conducted recently have corrected the methodological problems of earlier explorations. Since the 1990s, most of the SLA L2 writing studies have been conducted with pretest-posttest-delayed posttest designs which include control groups, and most have had a clear focus, with systematic and controlled feedback on a limited number of linguistic forms. In the current review, only those studies that include control groups and examine accuracy on new pieces of writing have been included.

Several studies exploring the effects of feedback on language acquisition have focused primarily on the degree of explicitness or the directness of the error correction.

Written CF is explicit in the sense that it informs the learner that he or she has made an error, but there are varying degrees of explicitness. In some cases, the correct form is given to the learner, generally termed “direct feedback”, whereas in other cases, the error is simply pointed out via underlining, circling, coding or other methods and is considered “indirect”. In alternative cases, the correct form is given directly, along with metalinguistic information or grammatical rules. In addition, decisions must be made on a continuum between narrowly focused (intensive correction of one or a limited number of errors) or unfocused (comprehensive) corrective feedback (the teacher corrects all the errors). The following sections will review the current findings concerning the effects of various forms of direct and indirect feedback as reported in comprehensive and focused written corrective feedback studies, respectively.

2.1.1 Focused and unfocused written corrective feedback

Only a few recent studies have investigated unfocused written corrective feedback, which provides comprehensive correction of every error in a student’s writing. Truscott and Hsu (2008) compared the writing performance of 47 ESL learners divided into two groups, one that received indirect comprehensive feedback, and another that served as the control group. Participants in the feedback group had their errors underlined. They outperformed the control group, who did not receive any form of feedback when it came to revisions; however, there was no significant difference in accuracy on a new piece of writing one week later. The authors, therefore, concluded that no learning had occurred.

In contrast, in a study with 62 randomly assigned secondary school L2 learners of Dutch, Van Beuningen, de Jong, and Kuiken (2008) found evidence that comprehensive feedback could lead to gains in grammatical accuracy on new pieces of writing. The

researchers compared two treatment groups who received comprehensive or unfocused feedback, one with direct and one with indirect feedback, where errors were underlined and a code was given (e.g., s = spelling error), and two control groups, one who practiced writing with no revisions, and one who self-corrected without the provision of feedback. Both the direct and indirect forms of comprehensive feedback yielded significant short-term effects, meaning that learners were able to revise their writing successfully, when compared to the two control groups. However, only the direct feedback group maintained accuracy in a subsequent writing.

Van Beuningen, de Jong, and Kuiken (2012) replicated the above study with 268 participants. As in 2008, the design included a pretest, treatment session, posttest, and a delayed posttest. This larger study likewise found that both direct and indirect forms of feedback allowed learners to revise a text effectively, i.e., short-term gains. This research was also the first to show that both direct and indirect forms of unfocused corrective feedback can lead to learning, in the sense of making fewer errors in new pieces of writing. The gains in accuracy for both treatment groups were durable and remained significant on the delayed posttest, which was given four weeks after the treatment. Interestingly, in this study a separate analysis of error types showed that direct feedback led to greater long-term gains in grammatical accuracy, whereas indirect feedback was more beneficial long-term for nongrammatical (i.e., lexical and spelling) errors. The authors suggest that this finding may explain the lack of effect for indirect feedback in Truscott and Hsu's (2008) comprehensive CF study, since the errors targeted in that study were for the most part grammatical in nature. Nonetheless, the conflicting results of these studies (Truscott & Hsu, 2008; Van Beuningen et al., 2008, 2012) indicate that more

research is needed before robust claims can be made about the effectiveness of direct and indirect forms of unfocused corrective feedback.

Very few studies have systematically compared the effects of focused and unfocused (comprehensive) corrective feedback. In a study involving Japanese EFL students in a university setting, Ellis, Sheen, Murakami, and Takashima (2008) compared the effects of focused (one linguistic target – article errors) versus unfocused (articles along with other errors - many linguistic targets) written corrective feedback and found significant and equal gains for both groups that received feedback in comparison to a control group that did not receive any feedback. This study offers evidence that both focused and unfocused written CF can be effective for English articles.

In another study, exploring the acquisition of English articles, however, Sheen, Wright, and Moldawa (2009) compared direct focused CF, direct unfocused CF and writing practice alone. The direct focused group only received feedback on articles, whereas the unfocused group was provided with feedback on a range of linguistic errors, including articles, copula ‘be’, regular past tense, irregular past tense and prepositions. Both treatment groups outperformed the control group in grammatical accuracy over time in all of the posttests. The focused group achieved the highest accuracy scores, not only for articles but, surprisingly, also for the other four structures that were tallied (articles, copula ‘be’, regular past tense, irregular past tense and prepositions), even though they did not receive feedback on those structures. The authors contend that the results provide evidence that unfocused corrective feedback is of limited pedagogical value, whereas focused corrective feedback can contribute to learning and improved grammatical accuracy in L2 writing. However, the superior performance of the focused group on the other structures, which were not corrected, is counterintuitive. The authors acknowledge

that the unfocused group received unsystematic treatment of errors, some being corrected and others not, which might explain this finding. This flaw in the execution of the study likewise calls into question the conclusion that unfocused feedback is of little value.

More recently, using a quasi-experimental design with university-level Chinese learners of English, Frear (2012) investigated the differential effects of focused direct feedback, unfocused direct feedback, focused indirect feedback, and unfocused indirect feedback on accurate use of English regular and irregular verb forms. Learners completed five narrative writing tasks and a revision task over a seven-week period. Feedback was provided on two of the writing tasks. Frear found improvements in accuracy on new texts for the regular past tense but not the irregular past tense. Compared to a control group, all four feedback groups showed short-term improvements, but only the focused direct feedback group showed improvements on a delayed written posttest suggesting that focused feedback is more durable than comprehensive feedback.

Finally, a meta-analysis recently conducted by Kang and Han (2015) revealed no significant differences between focused or unfocused feedback; however, the effect sizes for focused feedback were descriptively higher than those for unfocused feedback.

Clearly additional systematic inquiry is needed to determine whether or not comprehensive feedback, which is commonly used by classroom teachers, can improve linguistic accuracy for various grammatical structures. It may be that some structures can be treated with unfocused feedback, yet others (e.g., those that lack salience) will require more focused attention. Bitchener and Knoch (2015) call for replication studies of the one conducted by Van Beuningen et al. (2012) with adult learners, in instructed second language as well as foreign language contexts, in an effort to confirm the generalizability of the findings.

The jury is still out then on what is the most effective approach: focused or unfocused written corrective feedback, for helping learners acquire L2 forms and structures. Ferris (2010) concludes tentatively: “It seems fair to say that 1 or 2 features are too few and 15-20 features are probably too many” (p. 196).

2.1.2 Types of focused written corrective feedback

A number of recent studies have investigated the effectiveness of different types of focused corrective feedback, which targets only one problem or a very limited number of errors at a time. Bitchener, Young, and Cameron (2005) explored the effects of different types of focused direct corrective feedback on adult migrant ESL students by comparing explicit written feedback that included student-researcher five-minute conferences to explicit written feedback only. A control group that did not receive corrective feedback was also included. These researchers explored three types of errors (prepositions, the simple past tense, and the definite article) and found a significant effect for accuracy in new pieces of writing for the group that received both written and conference feedback for simple past tense and articles, but no overall effect when the three types of error categories were considered as one group.

In a later ESL study in 2008 with 75 learners, however, Bitchener focused solely on definite and indefinite articles and found that those who received written corrective feedback outperformed the control group and retained this level of performance two months later. In this study, three feedback combinations were tested (1) direct error correction with written and oral metalinguistic explanation (2) direct error correction with written metalinguistic explanation and (3) direct error correction; they were compared to a control group who did not receive any corrective feedback. Groups one and three

outperformed group two (just barely) and the control group. However, the study was later extended (Bitchener and Knoch, 2008) to include 69 additional learners, after which no differences were observed among the three treatment groups.

In one of the few longitudinal studies to examine direct written corrective feedback, Bitchener and Knoch (2010a) examined the effects of three types of direct focused feedback on the improved accuracy or use of two functions of English articles. Fifty-two low-intermediate level learners of English in New Zealand produced five pieces of writing. The first group received direct error correction with oral metalinguistic information in the form of a 30-minute lesson that included an opportunity for practice with the form. Group two received direct error correction with a written linguistic explanation; group three received direct feedback only, and group four served as the control and was not provided with any feedback. All three treatment groups outperformed the control group on all four posttests (one immediate and three delayed). The significant improvement in accuracy found in the feedback groups, in comparison to the control group, continued even after 10 months and offers compelling evidence of the beneficial role of corrective feedback. Differences in effects of the three types of feedback were not statistically different, indicating that the provision of direct correction alone was sufficient to produce durable gains.

Many of the studies conducted so far have focused on lower proficiency writers. A study designed to investigate the effects of written CF on two uses of the English article system, however, featured 63 advanced learners who already had relatively high levels of accuracy. Bitchener and Knoch (2010b) found that all three treatment groups (metalinguistic, metalinguistic with oral explanation, and indirect) outperformed the

control group on the immediate posttest. However, only the metalinguistic feedback groups continued this superior performance on the delayed posttest ten weeks later.

Sheen has also conducted several studies exploring the effects of both oral and written corrective feedback for adult intermediate students in the ESL classroom (e.g. 2007, 2010). Only one linguistic feature was targeted, English articles. In her 2007 study, she also explored aptitude as an additional factor in acquisition. She found that both groups that received corrective feedback (direct feedback and direct feedback with metalinguistic information) outperformed the control group that did not receive feedback, and that the direct metalinguistic group outperformed the direct only group in the delayed posttest. LAA was also shown to play an important role in acquisition, especially for the metalinguistic group. In a similar study in 2010, Sheen compared implicit oral feedback and two forms of explicit written feedback. Her findings suggest that the implicit oral recasts were less effective than the focused direct CF, which played a role in improving the learners' grammatical accuracy of English articles regardless of LAA (Sheen, 2010). She concluded that the degree of explicitness of corrective feedback appears to be more important than the medium, oral or written.

In a recent quasi-experimental study with 57 native speakers of Arabic at an American University in Lebanon, Diab (2015) examined the effects of direct correction along with metalinguistic feedback, metalinguistic feedback alone, and self-correction (control) on pronoun agreement errors and lexical errors. All groups received two class periods of form-focused instruction on pronoun agreement and lexical terms prior to the feedback sessions. Following instruction, learners completed three writing tasks in class and were asked to correct only their pronoun agreement and wrong word errors. Only the students in the two treatment groups received feedback to help them correct their errors.

At the time of the immediate posttest, all three groups (including the control) had reduced the number of pronoun agreement errors, but only the direct metalinguistic group had significantly reduced their number. By the delayed posttest, however, there was no significant difference among groups on pronoun errors. Concerning lexical errors, all three groups reduced the number of lexical errors at the time of the posttest, but there was no significant difference in error reduction for any of the groups, but a significant decrease in lexical errors was found for the direct metalinguistic group at the time of the delayed posttest. Since all three groups received form-focused instruction on the target structures prior to the feedback sessions, and since all three groups reduced both types of errors, it is hard to know whether or not the significant changes made by the learners who received direct feedback with metalinguistic information were really due to the feedback or other factors such as individual or class differences.

In a classroom-based (EFL) study with 89 Greek high-school learners of English, Stefanou and Révész (2015) explored two aspects of English article use, generic and specific plural reference. Participants were randomly assigned to one of three groups: direct feedback, direct feedback with metalinguistic comments, and a comparison group. Stefanou and Révész found an advantage for direct feedback as compared to no feedback, but no benefit for providing metalinguistic information in addition to direct feedback. Like Sheen's (2007), this study also investigated individual differences in grammatical sensitivity and further included a measure of meta-language knowledge. In contrast to Sheen, participants with greater grammatical sensitivity and knowledge of metalanguage were more likely to achieve gains in the direct feedback only group. There was no link found between individual differences and the gains made by the group who received direct feedback with metalinguistic comments. These results also concur with those of

Bitchener (2008) and Bitchener and Knoch (2008) who found no benefit for adding metalinguistic information in contrast to Sheen's (2007) findings that the provision of metalinguistic information led to superior gains.

Contrary to studies which found advantages for various forms of direct written corrective feedback, Shintani and Ellis (2013) found metalinguistic explanation alone to be more beneficial than direct feedback. In a study with 49 low-intermediate ESL students, they compared the effects of direct corrective feedback to those of metalinguistic explanation on the development of accurate use of English articles. The direct feedback had no effect on the development of accuracy in a new piece of writing. The metalinguistic explanation, however, led to gains on the immediate posttest, but the gains were not maintained on the delayed posttest. From these results, in conjunction with data they obtained from self-reports and eye-tracking, Shintani and Ellis proposed that the direct feedback did not benefit implicit or explicit knowledge, whereas metalinguistic explanation seemed to have enabled some development of explicit knowledge.

However, in a partial replication of the 2013 study, Shintani, Ellis, and Suzuki (2014) compared direct corrective feedback (DCF) to metalinguistic explanation (ME) on accuracy in the development of a more complex structure, the hypothetical conditional, as well as the indefinite article, and obtained conflicting results. In this study, 214 pre-intermediate level university learners of English in Japan were randomly assigned to five groups (1) ME, (2), ME with revision, (3) DCF, (4) DCF with revision, and (5) control. Feedback was found to improve accuracy in using the hypothetical conditional but not the indefinite article, and overall, the direct feedback was more effective and durable than metalinguistic explanation. Only the DCF with revision group maintained significant

long-term gains. Exit interviews led the researchers to speculate that the processing demands of attending to two grammatical forms at the same time may have pushed the learners to attend to the hypothetical, which was more crucial for the meaning and content of the texts. In sum, this study lends further support for the long-term benefits of direct corrective feedback and suggests that more complex forms are treatable with direct feedback.

Some L2 writing researchers have used depth of processing to suggest that a more indirect form of written CF (the error is pointed out but the correct form is not provided) is more effective than direct CF (the correct form and/or metalinguistic information are provided) because the former causes the learner to become engaged in the learning process (Ellis, 2010). However, it may be, as suggested by Ferris (2010), that this notion that learners should be left to determine the correct form on their own has more to do with the desire to teach students effective editing skills and less to do with language acquisition. Arguments in favor of direct forms of feedback suggest that it reduces confusion and provides information to sort out more complex (e.g., syntactic) errors (Bitchener & Ferris, 2012). The state of a learner's knowledge, however, may be a more critical factor (DeKeyser, p.c.) in determining the effectiveness of different forms of feedback. If a learner doesn't have clear declarative knowledge, then direct examples of form in conjunction with metalinguistic explanation might be required. On the other hand, if the learner has declarative knowledge, direct feedback that only supplies the correct form or even indirect feedback might suffice.

If the learner doesn't "notice" that he or she has made an error, then the learner cannot attend to it. Schmidt's noticing hypothesis (2001) offers support for the claim that in order for features of the input to be processed for acquisition, learners must first notice

them, but “noticing is therefore the first step in language building, not the end of the process” (Schmidt, 2001, p. 31). Schmidt (2001, 2010) also clearly makes a distinction between ‘noticing’ and ‘metalinguistic awareness’ by arguing that metalinguistic reflection requires a deeper level of thought than noticing. The question, then, is what is the best approach for helping learners not only to notice errors but to have accurate declarative knowledge essential for the development of proceduralized knowledge. It could be that direct written feedback would only lead to noticing, and that direct feedback with metalinguistic information would increase noticing and at the same time improve comprehension, which would result in a deeper level of cognitive processing.

In summary, studies which have compared different types of focused written corrective feedback have obtained mixed results. Bitchener et al. (2005) and Sheen (2007) found an advantage for direct error correction that included metalinguistic-information; Shintani and Ellis (2013) found an advantage for metalinguistic explanation; and Shintani et al. (2014) showed an advantage for direct error correction; however, other studies, such as Bitchener (2008), Bitchener and Knoch (2008; 2010a) and Stefanou and Révész (2015), found little or no difference between direct error correction and direct error correction that included metalinguistic information. Additional studies which investigate different structures and examine direct correction and the provision of metalinguistic explanation as separate variables are needed to sort out the long-term benefits of each.

2.1.3 Revision or provision?

There is also scant evidence as to what students should do with the feedback they are given. In some corrective feedback studies participants were required to edit their

errors (e.g., Ferris and Roberts, 2001; Truscott & Hsu, 2008; Van Beuningen et al., 2012), yet in other studies the students were just given back the corrected text to study, with no revisions required (e.g., Bitchener, 2008; Sheen, 2007). To date, only Shintani et al. (2014) have systematically compared the effects of feedback with and without revision. They found that requiring revision increased the effectiveness of the feedback.

The evidence overall suggests that revision after corrective feedback does indeed substantially help learners to reduce the number of errors on the piece of writing that has received correction; however, disagreement still exists concerning the long-term benefits of error correction on linguistic accuracy. Truscott (2007) maintains that although error correction may enable learners to improve the writing on which they received correction, there is no evidence that error correction leads to the acquisition of linguistic forms.

Other researchers, such as Sachs and Polio (2007), however, suggest that revision using CF plays a role in raising the learner's awareness of errors, which can lead to L2 development. Bitchener and Ferris speculate that

the cognitive processes involved in receiving and applying written CF in the revision of an existing text might yield the sorts of "gradual and nonlinear changes" in students' long-term language and writing ability that could also be hard to measure in short-term experimental studies of the "learning" facilitated by written CF, such as the ones conducted by Truscott and Hsu (2008). (2012, p.85)

Very few studies have specifically explored written CF followed by revision in a longitudinal study to examine the long-term effects on language development. In 2003, Chandler compared an experimental group that was required to revise marked errors after every first draft, and a control group. Five papers were written during the course of the semester. The experimental group significantly decreased its error rate over time compared to the control group. There was, however, no random assignment in this study, making it difficult to argue that the improvements were strictly a result of CF.

In another longitudinal study, using qualitative methodology, Ferris, Liu, Senna, and Sinha (2013) explored the improvement of 10 ESL learners who wrote four texts and were corrected on four error patterns. These researchers found great individual variation among students. All learners showed some improvement with at least some of the targeted errors; however, their progress was inconsistent.

In summary, further systematic research, exploring a variety of linguistic targets, is needed to determine the long-term effects of revision on linguistic accuracy.

2.1.4 Linguistic form

Beyond the number of errors to correct and the type of feedback to provide, there remains the question of which forms to correct. To date, the evidence suggests that written CF can help learners to acquire simple, rule-governed forms and structures, such as the simple past tense and specific aspects of the English article system (e.g., Bitchener et al. 2005; Bitchener 2008; Sheen, 2007). Evidence from oral feedback studies suggests that other morphosyntactic structures, such as question forms, could likewise benefit from written CF (e.g., Mackey & Oliver, 2002; Mackey & Philp, 1998), and it may be that written CF is even more effective at targeting complex forms and structures than oral feedback (Sheen, 2010). To date, however, little written corrective feedback research exists to test this prediction.

As pointed out by Doughty and Williams (1998) in their discussion of choice of linguistic form, “not all grammatical structures are acquired in the same way” (p. 211). Instructors have traditionally chosen to focus on forms that are repeatedly problematic for a specific group of students (Doughty & Williams, 1998). Although such teacher intuition may seem like a logical starting point, the students may not be ready to learn a particular

form. There is considerable research to suggest that there are developmental sequences that cannot be altered and steps that cannot be skipped (Pienemann, 1998, 2005, 2015; Pienemann & Keßler, 2011, 2012), but the list of documented sequences remains inadequate for pedagogic planning (Ortega, 2009). Furthermore, the way in which sequences relate to each other within the grammar of an individual learner has received little attention. Ortega concludes that “it would be unadvisable to develop instructional curricula around the known sequences and processes of L2 development” (2009, p.100).

Truscott (1996) claims that written CF will not facilitate the acquisition of syntactic knowledge that is more complex than discrete rule-based grammatical forms. Some grammatical rules may be too abstract for the student to make use of them; some forms are more salient in the input than others. Thus some forms may require more explicit feedback, and some might be better served by less explicit feedback. Ferris (2006) explored the effects of written corrective feedback on 15 categories of error. She examined thousands of errors generated by 92 ESL student writers and found that short-term ability to revise and edit errors did not necessarily lead to long-term gains for all categories. The students in the study made the most gains in grammatical accuracy during the semester with verb tense and form despite the fact that they were less successful in correcting these errors during the revision process. In contrast, the learners were very successful at editing sentence structure errors, yet showed no improvement with sentence structure errors over time. Ferris suggests that these results may have to do with the type of feedback that was provided for each error type. The instructors generally provided direct feedback for the sentence structure errors and indirect feedback for the “verb” category errors. She hypothesizes that the indirect form of feedback may have been more beneficial over time by consistently calling this error to the learners’ attention, and

thereby eliciting the guided learning and problem solving processes endorsed by Lalande (1982, p.14). This interpretation, however, contradicts the findings from the majority of the studies that have compared direct and indirect forms of feedback and have found direct feedback to be more beneficial overall for long-term gains in accuracy.

To sum up, the research to date concerning the best type of feedback to offer students (e.g., comprehensive versus focused; direct versus indirect; with or without revision) is still largely inconclusive. Furthermore, some errors may be less amenable to corrective feedback than others. Structures that lack salience, for instance, may require additional personalized practice in conjunction with corrective feedback. Although the evidence suggests that written corrective feedback can lead to learning, additional research, especially longitudinal, which investigates the value of various forms of indirect, direct, and metalinguistic feedback is needed to determine best practices for providing written corrective feedback for different categories of linguistic error.

2.2 Individual Differences

While the importance of feedback and the explicitness of and type of feedback needed for acquisition are still areas of disagreement for SLA researchers, most would now agree that some focus on form is necessary for adult L2 learners in order for noticing to take place. Cognitive interactionist theories (e.g. Long, 1996, 2009) claim that a focus on form is important to help learners create form-meaning mappings. Skill acquisition and learning theories (e.g. DeKeyser, 1998, 2015) also see a role for feedback in that it enables learners to proceduralize and eventually automatize declarative knowledge. The increase in accuracy that appears in new pieces of writing after the provision of written CF suggests that learning has taken place, but great individual variation often exists

among learners. Individual difference variables have been identified but have not been sufficiently explored for their potential in determining the effectiveness of various types of written corrective feedback. In an ESL setting, differences in proficiency level, L1, motivation, and aptitude likely play a role in the effectiveness of different types of corrective feedback and may explain some of the contradictions and inconsistencies in the literature.

2.2.1 Proficiency level

Truscott (1996) has argued that it would be impractical to personalize CF to each learner's current level of L2 development. At this point, however, very little research has examined the effects of CF in relation to proficiency level. Ferris and Hedgcock (1998, 2004), among others, have speculated that students at lower levels of L2 proficiency may not have adequate linguistic knowledge to correct errors that have been pointed out by the teacher. They suggest that a combination of direct and indirect feedback may be necessary, depending on the type of error. DeKeyser (2015) suggests that feedback will not be effective for grammar that is not yet known. Focus on form can only lead to proceduralization "provided that the requisite declarative knowledge is at the disposal of the learner" (DeKeyser, 2015, p. 102).

A meta-analysis conducted by Kang and Han (2015) found via a comparison of effect sizes that beginners did not benefit from written corrective feedback. In fact, the effect of written CF was negative when given to beginners. Both advanced and intermediate learners did benefit, and advanced learners saw more benefit than intermediate learners. These results lend support to the idea that developmental readiness

should be considered when providing feedback (Pienemann, 1998, 2005, 2015).

Some confirmation for this view comes from the results of oral corrective feedback studies (Ammar & Spada, 2006; Mackey & Philp, 1998; Trofimovich, Ammar, & Gatbonton, 2007), which showed greater benefits from oral feedback for higher than for lower proficiency learners. Yet, the majority of written CF feedback studies (e.g., Sheen, 2007; Bitchener, 2008) have in fact been conducted with groups of low or intermediate-level learners and have found encouraging results concerning the long-term benefits for language development. Two written CF studies by Bitchener and Knoch (2010a, 2010b) offer insight into the effects of feedback for learners from different proficiency levels. Both studies found, as previously discussed, that direct forms of feedback were beneficial for long-term accuracy gains for two uses of the English article system, both with lower level (2010a) and advanced level learners (2010b). Further research is still needed to investigate the effects of direct feedback for other structures with learners of varying levels of proficiency. It may be that some structures are amenable to direct feedback regardless of proficiency level, whereas other structures are not.

2.2.2 L1

In an ESL context, the diversity of native languages that are usually present creates yet another source of differences between individual learners. Speakers of different L1s may follow different developmental paths and have more difficulty with some forms than with others (Gass & Selinker, 2001). For instance, ESL learners are more likely to have difficulty with particular features or forms that have no close

equivalent in their L1. In a study which analyzed the interlanguage of speakers from five different language groups, Master (1987) found, in the case of English article usage, that acquisition patterns for learners whose L1 contained an article system differed considerably from those whose L1s did not include such a system. Transfer mistakes may also occur when the system in the mother tongue is similar but not identical to English (Swan & Smith, 2001). Where possible, research design and analysis should take L1-L2 differences into consideration.

2.2.3 Aptitude and motivation

According to Dörnyei and Skehan (2003), foreign language aptitude and motivation are the two most important individual differences for predicting second language learning success. Motivation, a “multifaceted” construct, may indeed play an important role in the case of corrective feedback, and there seems to be renewed interest in examining the role of motivation in SLA. However, very few studies have examined the role of motivation in relation to written corrective feedback. Goldstein (2006), for example, in a case study of two different L2 writers, concluded that the students’ motivation was a key factor that influenced the use they made of the teacher’s written feedback. More recently, Hyland (2011) examined learners’ attitudes towards written corrective feedback and found that motivation to use feedback was largely influenced by the students’ learning goals.

Aptitude is an equally important construct when considering corrective feedback or focus on form in SLA. It is generally agreed that LAA may predict the “rate of progress” for individual learners (Dörnyei & Skehan, 2003; Dörnyei, 2005). Evidence furthermore indicates that learners with a LAA will have greater success when making

the cognitive comparisons that are required for corrective feedback to result in learning (e.g., DeKeyser, 1993; Sheen, 2007; Shintani & Ellis, 2015). According to Dörnyei (2005), evidence shows that language aptitude is not restricted to certain teaching methodologies or settings, which suggests that high LAA may result in success for all forms of corrective feedback, although not necessarily to the same extent, regardless of how implicit or explicit the treatment.

Robinson (2001) offered a model of aptitude, in which cognitive resources and abilities are combined into aptitude complexes. One of the complexes Robinson proposes is *aptitude for explicit rule learning* (p. 374). His theory suggests that learners with high aptitude for explicit rule learning may benefit more from explicit explanation of linguistic features. Although he does not discuss written corrective feedback, we might expect metalinguistic written corrective feedback to prove beneficial to learners with high LAA. However, Robinson also suggests that metalinguistic abilities might be related to another complex, *aptitude for focus on form*, which allows learners to ‘notice the gap’. LAA, therefore, might also be necessary for direct feedback where learners have to generalize to new pieces of writing. Many studies have found a strong correlation between more implicit or inductive instructional conditions and higher language aptitude (e.g., Erlam; 2005, Robinson, 1997).

We can hypothesize, thus, that learners with greater LAA will do better than lower ability students in the case of less explicit corrective feedback, for example, where the error is either pointed out or marked with metalinguistic comments, yet the correct form is left for the student to determine. These same learners may also be better at internalizing metalinguistic information and applying it to new pieces of writing.

Dörnyei and Skehan (2003) suggest that there are also individual differences in “noticing abilities” and “pattern-extraction capacities” that allow some individuals to analyze and make generalizations more effectively than others (p. 599). A memory-oriented, less analytic learner, for example, might do better with corrective feedback chunks where only the correct form is provided in lieu of a more analytic form of CF where the rule is provided.

Only a few studies have systematically explored the role of individual differences in direct relation to corrective feedback. In a study that looked at error correction and oral proficiency, DeKeyser (1993) hypothesized that error correction would “widen the gap between the best and the weakest students in the class” (p. 504). Although his results were somewhat mixed, he did confirm that individual differences interact with error correction and that there were statistically positive effects for error correction if individual differences, such as previous achievement, motivation, and anxiety, were taken into account. In another study that examined oral corrective feedback, Havranek and Cesnik (2001) considered verbal intelligence, relative proficiency and attitude and found corrective feedback tended to benefit learners who had a higher LAA. Yilmaz (2013) found a strong relationship between grammatical inferencing and explicit oral feedback but no correlation between aptitude and the implicit feedback condition of recasts.

Concerning written CF, only three studies to date have considered LAA. Sheen (2007) found that a higher level of LAA correlated with the extent to which students benefited from both types of written CF tested in her study, which targeted one linguistic form, articles. She also found that learners with lower LAA were less likely to benefit from metalinguistic feedback.

Following up on Shintani et al. (2014), Shintani and Ellis (2015) examined the role of language analytical ability with different kinds of written feedback for two different structures, the past hypothetical conditional and indefinite articles. Like Sheen, Shintani and Ellis found that learners with stronger LAA benefited more from both types of feedback than learners with weaker ability. Second, LAA played a larger role for those who revised their original writing following the feedback. Third, the mediating effect was only evident in new writing produced shortly after the feedback. These researchers suggest that the extent to which LAA plays a role involves a complex interaction between type of feedback, opportunity to revise, and the target structure. For English articles, LAA was more strongly correlated in the direct condition when learners did not need to revise, yet LAA was more strongly implicated in the metalinguistic group when learners were asked to revise. In contrast, for the hypothetical conditional, the pattern was somewhat reversed. In the direct feedback condition, revision led learners to draw more on their LAA, whereas in the direct condition where no revision was required, it did not. These researchers speculate that LAA is more important when deep explicit processing is involved. Overall, their results do not lend clear support to the claim that LAA will play a stronger role when the feedback is metalinguistic in nature, since LAA benefitted learners who received either type of feedback.

Finally, Stefanou and Révész (2015) explored two aspects of English article use, generic and specific plural reference, and like Sheen (2007) this study also investigated individual differences in grammatical sensitivity. This study also included a measure of knowledge of meta-language. In contrast to Sheen, participants with greater grammatical sensitivity and knowledge of metalanguage were more likely to achieve gains in the

direct feedback only group. There was no link found between individual differences and the gains made by the group who received direct feedback with metalinguistic comments.

Thus, the role that LAA plays in written corrective feedback is unclear. Recently, SLA research has begun to consider individual subcomponents of language aptitude (e.g., DeKeyser & Koeth, 2011; Dörnyei, 2005; Kormos, 2013) and how these language aptitude components can be matched to instruction to optimize language learning (e.g., Doughty, 2013; Vatz et al., 2013). Aptitude research has also sought to determine which components predict high-level attainment. Based on a series of analyses conducted to determine whether or not the cognitive and perceptual abilities measured by the High-Level Language Aptitude Battery (Hi-LAB; Doughty et al., 2010) could distinguish very successful language learners from other individuals, Linck, Hughes, Campbell, Silbert, Tare, Jackson, Smith, Bunting, and Doughty (2013) determined that high-level attainment was related to associative memory, implicit learning, and phonological short-term memory. It may be, therefore, that finer grained explorations which consider individual components of language aptitude in relation to different types of written feedback are required.

In a review of the research that has examined how learner differences in aptitude interact with different types of second language instruction, Vatz, Tare, Jackson, and Doughty (2013) identify three possible interaction patterns. The first pattern is one in which the treatment “levels the playing field” and is best for all aptitudes. The second pattern is when one treatment produces superior results for all aptitudes, but that aptitude still plays a role, as appears to be the case in Sheen (2007). Finally, a third pattern finds both treatments to have similar outcomes, but one treatment’s effectiveness varies with aptitude.

Carefully controlled experiments on focused and unfocused written CF that systematically explore the various individual differences factors discussed above are relatively few. Consequently, much more research is needed to determine the role of these constructs in relation to error correction and to begin to identify possible patterns and interactions. The research reported here has attempted to offer some additional clues to the written CF puzzle by investigating the effects of written CF for learners of different levels of proficiency and to determine the effects of one individual difference, language aptitude. Overall, it is expected that a learner with higher LAA will outperform learners with lower LAA in all conditions of written corrective feedback.

Chapter 3: Research Questions and Hypotheses

Although recent studies have offered important insights into the effectiveness of written corrective feedback, many unanswered questions remain. The purpose of the present study was to contribute to the existing body of written CF research by examining whether or not there is a differential effect on EAP learners' acquisition of English grammatical forms depending on the type of corrective feedback given and whether or not aptitude plays a role in the learner's ability to make use of corrective feedback. Grammatical forms which to date have been underexplored were included to shed light on the efficacy of direct and metalinguistic feedback for different categories of error. By including a delayed posttest four weeks after the second treatment session, this study also aimed to shed more light on the durability of two different forms of written CF (direct and metalinguistic explanation). As discussed in the literature review, studies which have compared direct feedback to direct feedback which included metalinguistic explanation have yielded mixed results. Therefore, clear conclusions about the benefits of including metalinguistic explanation cannot be made until further investigations that separate these two aspects of the treatment have been conducted. Finally, the proposed research aimed to inform best practices for second language writing instructors who provide written corrective feedback for English as a Second Language (ESL) learners.

For the current study, in order to investigate the effects of direct corrective and metalinguistic feedback on different categories of linguistic error, four target structures were chosen based on the most common L1s currently represented at the research site: Spanish, Arabic, and Vietnamese. Errors with the simple past and present perfect were selected since "learning to use the tenses appropriately seems to be the greatest challenge

that ESL/EFL students face” (Celce-Murcia and Larsen-Freeman, 1999), and Arabic and Vietnamese learners of English, in particular, have difficulty with the present perfect (Swan & Smith, 2001). Pronoun dropping was selected as a target error as this is a persistent problem for Spanish learners of English, and can cause problems for learners from other pro-drop languages such as Japanese, Chinese, or Portuguese (Chang, 2001; Coe, 2001; Thompson, 2001). On the other hand, Arabic, Albanian, and Persian learners of English often produce resumptive pronouns, for example, after the subject in a relative clause (Swan & Smith, 2001). Similarly, topicalization is often seen in the writing of Spanish and Catalan speakers. Therefore, pronominal duplication was chosen as the third area of focus for this study (see Tables 1 and 2).

Table 1

Target Structures

Target Structures	L1s expected to have difficulty with target structure
errors with simple past and present perfect use	Arabic, Vietnamese
errors with simple past and present perfect forms (regular and irregular)	Spanish, Albanian, Arabic, Vietnamese
pronoun dropping	Spanish, Portuguese, Vietnamese, Japanese, Chinese
pronominal duplication	Arabic, Albanian, Persian, Spanish, Catalan

In addition to investigating different types of written CF and different categories of linguistic error, this study considered the interaction of LAA with different forms of corrective feedback. The research was designed to address the following research questions and hypotheses:

1. Does direct written corrective feedback (DCF) lead to more accurate use of the four targeted structures than writing with no feedback?

H1: The students who receive DCF will have greater gains in accuracy in the use of all four target structures than the control group who does not receive any form of written corrective feedback.

2. Does metalinguistic feedback (MF) lead to more accurate use of all four target structures than writing with no feedback?

H2: The students who receive MF will have greater gains in accuracy in the use of all four target structures than the control group who does not receive any form of written corrective feedback.

Support for the first two hypotheses came from several carefully designed studies which have reported that learners can improve their accurate use of targeted linguistic structures and forms via the provision of written corrective feedback (e.g., Bitchener et al. 2005; Bitchener, 2008; Sheen, 2007; Shintani et al., 2014; Van Beuningen et al., 2012).

3. To what extent is DCF effective in treating different error types?

H3a: The provision of DCF will lead to greater gains in accuracy with regular simple past and perfect forms than with irregular simple past and perfect forms.

H3b: The provision of DCF will lead to greater gains in accuracy with the use of the simple past than with the use of the present perfect.

H3c: The provision DCF will be equally effective in treating errors with pronoun dropping and errors with pronominal duplication.

Ferris suggests that errors with verb tense and form as well as problems with pronouns are treatable since they occur in a “patterned, rule-governed way” (1999, p. 6), and research on oral corrective feedback as well as studies on written corrective feedback

lend support for this claim (e.g., Bitchener et al., 2005; Han, 2002; Yang & Lyster, 2010). For instance, Frear (2012) found that written corrective feedback improved learners' use of the regular past tense, but not the irregular past tense. This finding supports the idea that structures for which there are patterned rules (regular past tense) are more treatable than for those for which there are no clear patterns. However, other factors, such as the salience of a feature and its relevance for meaning, play a role in error treatability. As put forth by DeKeyser (2005, in press), the complexity of an L2 structure and the difficulty it presents for learners can stem from a combination of factors including abstractness or novelty in meaning, the number of choices involved in choosing the correct morphemes for each form, and lack of transparency in form-meaning mapping. Thus, errors with verb tense choice (e.g., simple present vs. simple past) may be more treatable than errors which involve an understanding of aspect, which is more abstract (e.g., Ayoun, 2001; Ayoun, 2004; Ishida, 2004). Since learners' in the current study were expected to vary greatly concerning proficiency level, corrective feedback was expected to be less effective overall for the present perfect which is acquired later than the simple past (Bardovi-Harlig, 1997). Bardovi-Harlig (1997) did find that regardless of a learners' acquisition stage, instruction increased production of the present perfect tense, but that the intensity of the effect depended on the learners' stage of acquisition. Thus, both errors with verb tense were expected to respond to feedback; however, a fine-grained analysis was expected to reveal that irregular past tense forms, and the present perfect tense were less treatable than errors with regular past tense forms and the use of the simple past tense.

In the case of pronouns, the forms and meanings are not particularly complex; however, for learners with pro-drop L1s, the redundancy and optionality of pronouns may

reduce transparency (DeKeyser, 2005), which could diminish the effectiveness of written CF. Shintani, Ellis, and Suzuki (2014), for instance, speculate that learners paid more attention to a complex form, the hypothetical conditional, than indefinite articles, because the conditional form was more essential for meaning. It is possible then that learners who are given feedback on both pronouns and verb tenses may choose to focus more on verb tense, as this structure is important for meaning, and erroneous or dropped pronouns lack salience.

4. To what extent is the provision of MF effective in treating different categories of linguistic error?

H4a: The provision of MF will lead to greater gains in accuracy with regular simple past and perfect forms than with irregular simple past and perfect forms.

H4b: The provision of MF will lead to greater gains in accuracy with the use of the simple past than with the use of the present perfect.

H4c: The provision of MF will be equally effective in treating errors with pronoun dropping and pronominal duplication.

As with direct feedback, all of the targeted errors were expected to be treatable to some extent by written corrective feedback in the form of metalinguistic information. Since pronouns may be less salient and redundant, the provision of simple rules provided via metalinguistic explanation may be beneficial for students who lack declarative knowledge of these rules. However, in the case of irregular verb forms, metalinguistic feedback may be less effective as there are no rules (e.g., Frear, 2012). Bitchener et al. (2005) corrected errors with the simple past tense and found that errors in choice of tense

were amenable to direct explicit written feedback (the correct tense to use was indicated but the form was not given), but in this study, the group which received direct correction along with conferences (during which errors were explained and additional examples were provided to explain the form) significantly outperformed the group which only received direct corrective feedback. This finding suggests that in the case of verb use, both examples of the form and metalinguistic information may be necessary. Other studies have shown that the provision of metalinguistic information is beneficial for both simple and complex forms (e.g., Shintani et al., 2014). Yet, Ayoun (2004) suggests that “subtle and complex aspectual distinctions” may require both the explicit positive feedback that comes with metalinguistic instruction and the negative feedback that comes from recasts. In summary, metalinguistic feedback was expected to result in an improvement in accuracy for all targeted errors, but errors with regular simple past and errors with pronoun usage were expected to be more treatable than errors which require an understanding of both tense and aspect.

5. Is there any difference in the effect of DCF and MF on learners’ accurate use of the four grammatical features?

H5a: The provision of DCF will have a greater effect on learners’ accurate use of the simple past and the present perfect than the provision of MF.

H5b: The provision of DCF will have a greater effect on learners’ accurate use of pronouns (inclusion and exclusion) than the provision of MF.

Support for these hypotheses came from studies that have compared direct corrective feedback alone to direct feedback which included metalinguistic explanation (Bitchener and Knoch 2010a; Bitchener and Knoch 2010b) and have found little difference in long-term accuracy gains between groups who were provided with direct

written CF and those who were provided with direct written CF with metalinguistic explanation. In the only study to date to compare the provision of DCF to the provision of ME that did not include DCF, Shintani et al. (2014) found greater long-term gains for direct feedback, even in the case of a complex rule, than for metalinguistic information. However, in their study the metalinguistic group was given instruction on the rule as a group and had to identify their own errors in compositions. In the current study, the metalinguistic explanation was individualized and provided in response to specific errors in the same way that the direct feedback was provided. Hypotheses 5a and 5b are also supported by the lion's share of the studies that have compared direct and indirect forms of feedback and have found direct forms of feedback to be more beneficial overall for long-term gains in accuracy.

6. To what extent does the learners' language-analytic ability determine the effectiveness of DCF or MF?

H6a: Learners with a higher language-analytic ability will have greater overall gains in accuracy for both CF treatment groups.

H6b: Learners with higher language-analytic ability will benefit more from the metalinguistic feedback than learners with lower language-analytic ability.

This hypothesis is based on the premise that metalinguistic explanation requires greater depth of processing by learners and could lead to explicit knowledge that will enable learners to improve their accuracy on new pieces of writing.

H6c: Learners with lower language-analytic ability will benefit more from the DCF than they will from the MF.

Support for these hypotheses came from Sheen (2007)¹, who found a differential effect of aptitude for different forms of feedback. Learners with lower LAA were less likely to benefit from metalinguistic feedback.

¹ The studies by Stefanou and Révész (2015) and Shintani and Ellis (2015) that likewise explored interactions between language aptitude and written corrective feedback were not published at the time these hypotheses were written.

Chapter 4: Methodology

4.1 Setting and Participants

The participants for this study were recruited from 14 English for Academic Purposes (EAP) classes (low-intermediate to advanced levels²) at the St. Petersburg Gibbs, Clearwater, Seminole, and Tarpon Springs campuses of St. Petersburg College in Florida. Students enrolled in the English for Academic Purposes program strive to improve their level of English in order to matriculate into regular college-level courses, and they are accustomed to writing and revising several essays during the course of a semester. Learners in this program have very diverse L1s; however, three native language groups, Spanish, Arabic, and Vietnamese, were well represented at the time of this study and provided a basis for the target structure selection (discussed in the next section).

The current program offers five levels and follows a content-based integrated-skills approach. Students from the four highest levels of EAP writing courses were asked to participate in the study and were randomly assigned to one of three groups: the direct feedback group, the metalinguistic feedback group, and the control group. The lowest level course was not targeted as learners in this level were unlikely to have the proficiency needed to complete all parts of the data collection.

One hundred and sixty-five participants consented to participate in the study. One participant requested to be excused from the study after the first writing task. Three participants withdrew from the EAP program during the course of the semester and

² The EAP program at St. Petersburg College offers five levels of integrated-skills reading, writing, and grammar courses. Participants for the study were recruited from levels two through five. Learners are placed into the program based on the Levels of English Proficiency Test (LOEP), part of College Board's ACCUPLACER Program. Learners in levels two through five have LOEP Language Use and Reading Skills scores which range from 80 to 120. Writing samples are also used to determine placement. Some participants may have been in their first semester of study, whereas others may have already completed two or three semesters of EAP.

therefore did not complete the study. Nine others failed to complete all parts of the protocol due to repeated absences during the semester. Thus, only one hundred and fifty-one participants completed all parts of the data collection (administered in six sessions). There were 27 different L1s, with substantial groups of native Spanish speakers ($N = 59$), Vietnamese speakers ($N = 28$), and Arabic speakers ($N = 23$). See Table 2 for a detailed list of native languages for those participants who completed all six sessions. The 151 participants who completed all sessions ranged in age from 18 to 61 years old with a mean age of 29.80 ($SD = 10.91$). There were 59 males and 92 females.

Table 2

Participants' Native Languages

Native Language	<i>N</i>	Direct	Metalinguistic	Control
Spanish	59	18	20	21
Vietnamese	28	10	8	10
Arabic	23	9	8	6
Albanian	4	1	1	2
Russian	4	2	1	1
Chinese	3	1	1	1
Portuguese	3	1	1	1
Ukrainian	3	1	2	0
Farsi	2	1	0	1
German	2	1	1	0
Greek	2	1	1	0
Korean	2	0	1	1
Thai	2	1	0	1
Bulgarian	1	0	0	1
Catalan	1	1	0	0
Creole	1	1	0	0
Croatian	1	0	1	0
Filipino	1	0	1	0
Grebo	1	0	0	1
Hebrew	1	0	1	0
Italian	1	1	0	0
Khmer	1	0	1	0
Kyrgyz	1	0	1	0
Latvian	1	0	0	1
Norwegian	1	0	0	1
Punjabi	1	1	0	0
Tagalog	1	1	0	0

Responses to background questions asked via a survey administered at the last session revealed that nine learners in the study had been in the U.S. for less than six months, 23 learners had been in the U.S. for six months to one year, 47 had been here for one to three years, and 72 had been living in the U.S. for more than three years.

4.2 Target Structures

To date, the majority of the focused feedback studies (e.g., Bitchener 2008; Bitchener & Knoch, 2008, 2010a, 2010b, Ellis et al., 2008, Sheen, 2007, 2010; Sheen et al. 2009) have investigated the same structure, two specific functions of the English article system and found that groups who received corrective feedback outperformed the control group for these specific uses of English articles. Very few written CF studies have targeted other linguistic error domains and categories. One exception is the study by Bitchener et al. (2005), which investigated not only the use of the English article system, but also the simple past tense and prepositions. These researchers found that written CF helped learners to improve their accuracy with specific uses of articles and verb tense, but not with the more idiosyncratic uses of prepositions. Another exception is Shintani et al. (2014), who found that written CF was effective for improving accuracy with a syntactic structure, the hypothetical conditional. There remain, however, many structures which have received little or no focused research attention.

The current research, therefore, aimed to examine the effectiveness of different forms of corrective feedback on errors with two verb tenses, simple past and present perfect, and two errors with pronouns, specifically pronoun dropping and pronominal duplication.

Errors with simple past and present perfect tense were investigated since errors of use with verbs are often persistent regardless of the learner’s L1 and proficiency level. Comprehending the semantics of where the past tense ends and the present perfect begins is particularly tricky for ESL students (Celce-Murcia & Larsen-Freeman, 1999). It was anticipated that Arabic learners of English in particular would have difficulty with the use of the perfect tense since there is no distinction between actions accomplished in the past with and without a connection to the present in Arabic. Use of the present perfect tense was likewise expected to be problematic for other learners whose L1s have no equivalent tense to the present perfect tense such as Vietnamese, where tense and aspect are generally understood from the context (see Table 3 below and Table 1 in chapter 3).

Table 3

Expected Errors Based on L1

L1	Target Structures
Arabic	<ul style="list-style-type: none"> • resumptive pronouns • errors with simple past and present perfect
Spanish	<ul style="list-style-type: none"> • pronoun dropping • topicalization • errors with simple past and present perfect
Vietnamese	<ul style="list-style-type: none"> • errors with simple past and present perfect

Pronoun dropping was chosen as a target, since Spanish learners have a tendency to drop subject pronouns, particularly in subordinate clauses (e.g., *When Mary came to the door, *spoke to the salesman*). Learners from other pro-drop languages were likewise expected

to make this error (e.g., Chinese, Portuguese). Finally, pronominal duplication errors were selected, since resumptive pronouns are commonly provided by Arabic, Albanian, and Persian learners whose L1s permit this structure (e.g., The man who works at the store *he* is helpful), and errors with topicalization are sometimes seen in Spanish and Catalan speakers (e.g., Classrooms *they* have desks).

4.3 Treatment

Corrective feedback was provided electronically for errors that occurred during two in-class writing tasks with simple past and present perfect, and with pronoun dropping and pronominal duplication. Course instructors were asked to refrain from explicit teaching of the target structures during the course of the study. Learners were corrected when they made an error with simple past and present perfect verb forms (e.g., The man *has slept* in the street) and when they made an error with choice of verb tense (e.g., I *saw* a lot of movies lately), but only in cases where either simple past or present perfect was required.

The type of corrective feedback given varied for each treatment group. The direct-only correction group received the traditional strategy that consists of marking the error on the student's essay and indicating the correct form. The metalinguistic correction group likewise had their errors marked; this group received consistently worded metalinguistic comments (e.g., brief grammar rules) but did not receive the correct form. In the current study, metalinguistic comments were also considered explicit and direct in the sense that they were provided electronically within the text next to each error, similar to the metalinguistic feedback provided by Sheen (2007), which was given in addition to direct feedback. This was likewise the case for the metalinguistic feedback provided in

the studies conducted by Bitchener and Knoch (2010b) and Diab (2015), where the rules were provided on the page and errors were indicated with an asterisk or underlined. Other studies investigating metalinguistic feedback have taken a more indirect approach by giving metalinguistic information in the form of a lesson or handout where individual errors are not indicated on the student's writing (e.g., Shintani & Ellis, 2013; Shintani et al., 2014). The control group received general comments on content and organization, for ethical reasons, and also in an attempt to avoid a Hawthorne effect. The control group did not, however, receive any grammatical error correction.

4.4 Procedures

The current study employed a pretest-treatment-posttest-delayed posttest experimental design. One week prior to the first writing task, the students were given the LLAMA F, a sub-set of the LLAMA computer-based aptitude test (Meara, 2005), as a measure of LAA. The LLAMA F specifically measures grammatical inferencing or the ability to induce the rules of an unknown language (e.g., explicit inductive learning). At this time, participants were also given a form-focused pretest³ (30 items) to measure prior knowledge of the target structures (see Appendix A for test questions). Participants were matched in sets based on LLAMA F scores and L1 (see Table 2 for L1s by group), after which members of each set were randomly assigned to either one of the two treatment groups (direct corrective feedback or metalinguistic feedback) or the control group. LLAMA F scores, thus, ranged from 0 to 90 for all three groups. Mean scores and standard deviations for the LLAMA F are reported in Table 5.

³ The pretest was piloted prior to the study with approximately 15 learners from the same EAP program.

The following week, during one hour of class time, all three groups were asked to write an essay (Essay 1). The students used a word processor for the writing task and were instructed to write at least two pages using Times New Roman 12-point text. The directions were the same for all three groups, and all participants were instructed by the researcher to pay attention to accuracy (see Appendix B for the writing prompts). The students submitted their essays to a drop box within Desire To Learn, the learning management system (LMS) used by the college. The researcher provided all of the feedback via the LMS, which allowed for systematic and clear mark-up. The essays produced by the students were corrected within one week by the researcher, and depending upon group membership, the students received direct, metalinguistic, or only brief content and organizational feedback, as described above. All four error types were corrected for participants assigned to one of the experimental groups. Inevitably, each student produced different numbers of obligatory contexts for each structure; therefore, the amount of feedback provided differed from learner to learner. The researcher then returned to each classroom the following week (Week 3) and returned the essays with feedback (released via the LMS) to the students.

The students were given 30 minutes in class to revise their essays using the feedback. The revised essays were then collected. Learners were not given feedback on the quality of their revisions. The following week (Week 4), students wrote a second essay in class (Essay 2). The second piece of writing for all groups was likewise corrected and returned one week later (Week 5), at which time the students were again given 30 minutes to revise. After a break, but immediately following this revision session, learners were asked to write on a new topic (Essay 3, which served as the immediate posttest). An unannounced delayed written posttest (Essay 4) was conducted

four weeks later (Week 9), at which time the students were asked once again to write on a new topic. During this session, learners were also given a form-focused posttest on the targeted structures that followed the same format as the pretest. The items for the posttest were essentially the same as the items on the pretest with superficial changes (e.g., lexical elements and item order). Following the form-focused posttest, participants were asked to complete a brief background questionnaire, which requested information about the participants' age, gender, L1, educational background, and length of time spent in the United States. The questionnaire also asked a few questions to assess how students felt about their errors, how they felt about error correction, and how they viewed and utilized different types of feedback (see Appendix C).

Table 4

Experimental Procedure

Group Condition	Week 1	Week 2	Week 3	Week 4	Week 5	Week 9
Direct CF	Aptitude test and Pretest on target structures	Essay 1	Essay 1 revised by learners based on focused direct feedback	Essay 2	Essay 2 revised by learners based on focused direct feedback + immediate posttest Essay 3	Delayed Posttest Essay 4, Posttest on target structures, Exit questionnaire
Metalinguistic CF			Essay 1 revised by learners based on focused metalinguistic feedback		Essay 2 revised by learners based on focused metalinguistic feedback + immediate posttest Essay 3	
Control			Essay 1 revised by learners based on self-correction and general comments		Essay 2 revised by learners based on self-correction and general comments + immediate posttest Essay 3	

4.5 Writing Tasks

The writing prompts used for the study were comparable concerning presentation and design, and the order of the prompts (topics) was counterbalanced over individuals to control for differences in writing task difficulty. Students were asked to narrate a series of images and provided information on various themes (i.e., the life and achievements of Leo Messi, the life and achievements of Angelina Jolie, changes in communication over

the past 50 years, and changes in education over the past 50 years; see Appendix B for the writing prompts). Prior to the study, the prompts were piloted with approximately 15 students from the EAP program to verify that each would elicit the desired target structures and errors. Students were given one hour to write for each new prompt and were asked to write at least two pages, using double spacing and 12-point, Times New Roman font. To avoid potential individual differences in vocabulary knowledge, words for unfamiliar items in the pictures were provided on the board at the start of each writing session.

4.6 Operationalization of Variables

All written output produced by the participants was coded for correct and incorrect uses of each targeted linguistic feature. For each text, accurate use of the simple past and present perfect were tallied, and percentage correct was calculated based on the total number of obligatory contexts for each tense within a text. For pronoun dropping and pronominal duplication, all possible subjects within the text were tabulated along with the number of dropped and redundant pronouns.

The coding process was blind, in that the researcher coded without knowledge of which group each text belonged to. Participants were given random identification numbers, which were used in lieu of names on the essays during the coding process. A log was kept to track the numbers and group membership, but was not referenced during coding. To examine the reliability of the scoring, 76 texts (over 12 percent of the 604 texts) were rescored by a trained teacher/researcher, who has extensive experience as an instructor at the research site. Training on coding was conducted by the researcher using essays written during the pilot phase. The overall agreement rate between the two scorers

(both native speakers of English) was high for all four structures as determined by the intra-class correlation coefficient of .95 for the simple past tense, .89 for the present perfect tense, .82 for pronoun dropping, and .94 for redundant pronouns.

Chapter 5: Results

The first two subsections of this chapter report the descriptive statistics for the LLAMA F test of grammatical inferencing and the descriptive statistics for the three subsections of the form-focused grammar tests, along with the results of ANOVAs that were conducted to compare group pretest and posttest scores. The next four subsections report the results of the analyses conducted using the data obtained from the four essays written by each participant during the study. These results are presented separately for each target structure. The final subsection reports the results of the exit survey concerning feedback preferences.

5.1 Descriptive Statistics for LLAMA F

LLAMA F scores ranged from 0 to 90 for each group. Additional descriptive statistics for the LLAMA F are presented in Table 5. LLAMA F means for the three groups were compared using a one-way ANOVA. As expected, since participants were matched in sets based on LLAMA F scores and L1 before being randomly assigned, no significant difference was found between the three groups ($F(2, 148) = .151 p > .05$).

Table 5

Descriptive Statistics for LLAMA F

Group	<i>N</i>	<i>Mean</i>	<i>SD</i>
Direct	52	32.5	25.35
Metalinguistic	50	31.60	24.77
Control	49	29.80	25.21

5.2 Form-focused Grammar Test

Reliability estimates for the 30 item pretest and the 30 item posttest were calculated using the data from the 151 participants who completed all data collection sessions. Cronbach's alphas for the pretest and the posttest were .831 and .856, respectively. Each section of the form-focused grammar pretest and posttest was then analyzed separately. There were 10 questions in each section. The number of correct items in each section were tabulated to obtain each score. The descriptive statistics for each section are reported in Table 6.

Table 6

Descriptive Statistics for Form-focused Grammar Test (Mean with SD in Parentheses)

Group	N	Verb Tense	Pronominal Duplication	Pronoun Dropping
Direct	52			
Pretest		7.67 (2.07)	7.90 (2.55)	8.25 (2.12)
Posttest		6.25 (2.48)	8.33 (2.21)	8.50 (1.66)
Metalinguistic	50			
Pretest		7.90 (1.89)	8.72 (1.87)	8.36 (1.93)
Posttest		6.64 (2.35)	9.00 (1.44)	8.64 (1.56)
Control	49			
Pretest		7.65 (1.80)	8.41 (2.22)	8.08 (1.70)
Posttest		6.43 (2.42)	8.59 (2.40)	8.14 (2.28)

In order to determine whether or not there were differences between groups on the form-focused grammar test at the start of the study, a series of one-way ANOVAs were conducted to compare group means for each section of the test. For the pretest, no significant differences were found between groups for any of the sections: verb tense ($F(2, 148) = .253 p > .05$), pronominal duplication ($F(2, 148) = 1.733 p > .05$), and pronoun dropping ($F(2, 148) = .260 p > .05$). For the posttest there was still no significant

difference found between groups for any of the sections: verb tense ($F(2, 148) = .329 p > .05$), pronominal duplication ($F(2, 148) = 1.379 p > .05$), and pronoun dropping⁴ ($F(2, 148) = .942 p > .05$)⁵.

5.3 Simple Past Tense in the Writing Tasks

Simple past tense scores were analyzed to examine the effect of feedback type on accurate use of this verb tense in subsequent writing tasks. For this analysis, a total of 12 participants were eliminated from the data set. Eight were deleted who demonstrated command of the simple past tense. These learners did not make any errors with simple past on Essay 1 nor on Essay 2, and therefore, did not receive any feedback on this structure. Four additional participants failed to follow the writing prompts. By writing off-topic, these four learners had zero obligatory contexts for the simple past in at least two of the essays written for the study, causing their scores on these essays to become outliers. Furthermore, these four participants had zero attempts at the structure on either Essay 1 or Essay 2 or both, and therefore received no feedback, or only one round of feedback, as opposed to two rounds like the other participants. In short, 139 participants were included in the analysis for the simple past tense.

For the simple past data analysis, the outcome variable was accurate use of the simple past tense. For each essay, all obligatory instances of simple past were coded with a 1 when correct and a 0 when incorrect. Verbs were considered incorrect both when the wrong tense was chosen (e.g., present tense was used in lieu of simple past) and when

⁴ Errors with pronoun dropping and pronominal duplication only occurred with a small subset of participants, so the form-focused grammar test scores for these small subsets were likewise examined (see sections 5.5 and 5.6 for these results).

⁵ The form-focused pretest and posttest were expected to be of comparable difficulty; however, these tests were not equated prior to the study making their equivalence uncertain.

there was an error with form (e.g., caught instead of caught). The percentage correct for each essay for each learner was calculated to obtain accuracy scores.

The assumption of normality for simple past accuracy scores was not satisfied for any of the groups for any of the essays, as assessed by both Shapiro-Wilk's and Kolmogorov-Smirnov tests ($p < .05$) and visual inspection of the data. Therefore, Arcsine transformations were performed. Although the resulting scores were still not normally distributed, a series of ANOVAs and ANCOVAs were conducted considering that these procedures are generally robust to violations of normality. Unfortunately, there were other more problematic violations revealed during these tests. There was no homogeneity of variance, as assessed by Levene's test of homogeneity of variance ($p < .05$). There was likewise no homogeneity of covariances, as gauged by Box's test of equality of covariance matrices ($p = .001$).

Written corrective feedback studies have reported (e.g., Ferris et al., 2013) that there is great variation between individuals in their response to feedback, and certainly in the present study the participants, who came from different levels of EAP courses, varied as to their fluency, accuracy, and ability to improve over time. Moreover, the number of trials or attempts to use the simple past tense also varied not only between participants but from essay to essay for each learner. Thus, the percentage correct for each structure was estimated with a different number of trials for each participant; this does not improve the issue of homogeneity of variance, an important assumption for ANOVA. Mixed-effects logit models, however, allow every observation to be modeled while accounting for the fact that there are many observations for each participant (Jaeger, 2008). All data, therefore, were analyzed using mixed-effects logit models in R (R Development Core Team, 2011) with the R package *lme4* (Bates, Maechler, & Bolker, 2011).

The predictors in the model included group membership (direct, metalinguistic, or control), LLAMA F scores, and scores from the verb use section of the form-focused grammar pretest. To aid with model convergence and to put these predictors on the same scale, LLAMA F scores and the grammar pretest scores were standardized using the scale function in R to center scores on the mean. For the 139 participants included in this analysis, LLAMA F scores ranged from 0 to 90 with a mean of 31.3 and a median of 30. Grammar scores for the verb section of the pretest ranged from 3 to 10 with a mean of 7.851 and a median of 8. Participants were included as a random effect in the model. A total of 8,106 observations from 139 participants were included in the analysis (see Table 7 for a breakdown by group and essay time). The percentage correct for simple past tense for each group at each essay time is given in Table 8, and the number of correct and incorrect trials with the simple past tense at each essay time for each group are visually presented in Figure 1.

Table 7

Total Number of Obligatory Contexts for Simple Past Tense at Each Essay Time

	Essay 1	Essay 2	Essay 3	Essay 4	Total
Direct feedback group ($n = 47$)	558	866	815	735	2974
Metalinguistic feedback group ($n = 45$)	545	742	694	650	2631
Control group ($n = 47$)	574	638	664	625	2501
Total for all groups	1103	1608	1509	1385	8106

Table 8

Percentage Correct for Simple Past Tense at Each Essay Time

	Essay 1 % correct	Essay 2 % correct	Essay 3 % correct	Essay 4 % correct
Direct	67	73	80	78
Meta	78	83	85	78
Control	73	71	75	73

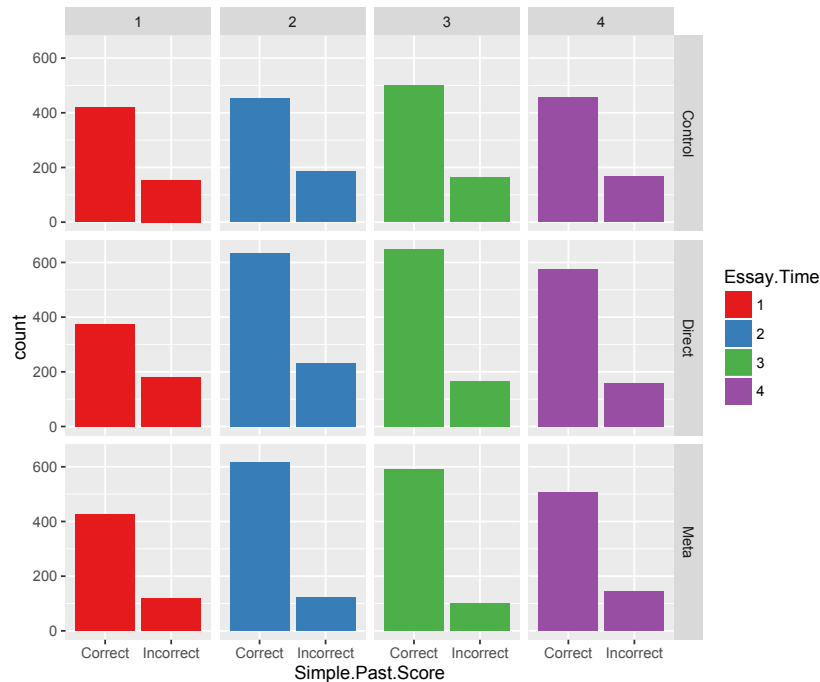


Figure 1. Correct and incorrect trials with simple past tense for each group at each essay time.

To examine the effect of group on accurate use of the simple past tense, a mixed-effects logit model with group, essay time, and the interaction between group and essay time as fixed effects, and participant as a random effect, was fit. Initial analyses included all four essay times. However, in order for the model to converge with the above interaction terms, it was necessary to eliminate one of the essay times from the model. Figure 1 revealed that for both treatment groups there were sizable gains in accuracy between essay time one and time two, after one round of feedback, but in fact, little change between time two and time three. Since the purpose of the study was to determine if two rounds of feedback would lead to improvements on a third writing task, the reported model includes only the written pretest (Essay 1), the immediate written posttest (Essay 3), and the delayed posttest (Essay 4).

A random intercept allows variance in starting points to be modeled for each participant, whereas random slopes allows the effect or rate of change to vary for each participant; therefore, given the large range in proficiency levels and individual differences, both were considered. The first model, which only included random intercept, was compared to one that included both random intercept and random slope. The results of a comparison between models showed that random slopes significantly improved model fit, $\chi^2(6) = 84.091, p < 0.001$.

Additional models, which included standardized LLAMA F scores and standardized form-focused pretest scores and interactions with these effects, were fit and compared. The best model, reported below and presented in Table 9, includes pretest scores, LLAMA F scores, group, essay time, and the interaction between group and essay time, as well as interactions between LLAMA F and group, LLAMA F and essay time, and LLAMA F, group, and essay time, as fixed effects. The model was likewise fit with random effects for participant, including the random intercept and random slopes for essay time. Because essay time was within-participant, but treatment group was between-participant, this random effect structure constitutes the “maximal” random effect structure possible (Barr, Levy, Scheepers & Tily, 2013).

Pretest scores were significant as a predictor in the model ($\beta = 0.18, SE = 0.09, p = .029 *$). These scores accounted for a significant amount of variance in accuracy; however, including this predictor did not substantially change any of the effects. The large standard deviation for participants for the random intercept (.74) and the standard deviations for random slopes seen at essay time three (.65) and time four (.77) confirm that there is indeed great variance among individuals in both starting point and gains over time.

Table 9

Result Summary of Coefficient Estimates β , Standard Errors SE, Associated Wald's z-score, and Significance Level p for the Fixed and Random Effects in the Simple Past Tense Analysis⁶

	Coef. β	SE	z	p	Participant (SD)
Intercept (Control)	1.25	0.19	6.58	0.000 ***	0.74
Pretest	0.18	0.08	2.18	0.029 *	
LLAMA	0.30	0.19	1.57	0.117	
Direct	-0.47	0.26	-1.78	0.075 .	
Metalinguistic	0.13	0.27	0.47	0.639	
Essay Time 3	0.07	0.21	0.38	0.708	0.65
Essay Time 4	-0.12	0.19	-0.60	0.546	0.77
LLAMA:Essay Time 3	-0.18	0.21	-0.87	0.387	
LLAMA:Essay Time 4	-0.08	0.20	-0.43	0.666	
LLAMA:Direct	-0.55	0.27	-2.06	0.039 *	
LLAMA:Metalinguistic	0.02	0.28	0.06	0.955	
Essay Time 3:Direct	0.79	0.28	2.79	0.005 **	
Essay Time 4:Direct	0.88	0.27	3.31	0.001 ***	
Essay Time 3:Meta	0.54	0.30	1.84	0.066 .	
Essay Time 4:Meta	0.19	0.27	0.69	0.490	
LLAMA:EssayTime3:Direct	0.60	0.29	2.07	0.038 *	
LLAMA:EssayTime4:Direct	0.19	0.27	0.70	0.487	
LLAMA:EssayTime3:Meta	-0.42	0.31	-1.34	0.180	
LLAMA:EssayTime4:Meta	-0.08	0.28	-0.27	0.786	

Note: * = $p < .05$; ** = $p < .01$; *** = $p < .001$

The model reported in Table 9 was used to assess several paired comparisons. At Essay Time 1, which served as the written pretest, the direct feedback group (69%) was

⁶ Calculating the correlation between random effects was not possible in the full model. When included, the model failed to converge. The reported model thus, does not estimate the correlation between random effects but rather sets it to zero. In a scaled-down model that only included Essay Time 1 and 3, it was possible to calculate the correlation between random effects. The correlation between the random intercept and the random slope of $-.52$ suggests that participants who started lower benefitted more than those who started higher.

significantly less accurate than the metalinguistic group (80%, $\beta = -0.59$, $SE = 0.27$, $p = .026$), and marginally lower than the control group (78%, $\beta = -0.47$, $SE = 0.26$, $p = .075$). The metalinguistic and control groups, however, were not significantly different from each other ($p = .639$). Since the groups began at different levels of predicted accuracy, the most important results concern the gains made by each group.

The results indicate that there was a significant interaction between essay time and group. The direct feedback group showed a significant increase in probability of a correct response from the time of the written pretest (69%) to the time of the immediate posttest (84%, $\beta = .87$, $SE = 0.20$, $p < .001$). These gains were maintained at the time of the delayed posttest (82%, $\beta = .77$, $SE = 0.19$, $p < .001$). As shown in Table 10, the direct feedback group is predicted to produce accurate simple past tense verbs 69% of the time on the pretest (Essay 1). By the time of the posttest (Essay 3), these learners are predicted to accurately use the simple past tense 84% of the time. In other words, they have reduced their error rate from 31% to only 16%, essentially a 50% reduction in error rate. At the time of the delayed posttest (Essay 4), the direct feedback group was still predicted to accurately use the simple past tense 82% of the time, meaning that the significant gains seen at on the immediate posttest were durable. The model also indicates that these gains in accuracy achieved by the direct feedback group were significantly different compared to changes seen at Time 3 ($\beta = -0.79$, $SE = 0.28$, $p < .005$) and Time 4 ($\beta = -0.88$, $SE = 0.27$, $p < .001$) in the control group.

The metalinguistic feedback group likewise made significant gains in accuracy from the time of the pretest (80%) to the time of the immediate posttest (88%, $\beta = .63$, $SE = 0.22$, $p = .005$). The model further indicates, however, that this change in accuracy was

only marginally different from changes seen in the control group ($\beta = -0.55$, $SE = 0.30$, $p < .066$). Furthermore, the metalinguistic group lost these gains by the time of the delayed posttest (81%, $\beta = .07$, $SE = 0.20$, $p = .718$). The metalinguistic group is predicted by this model (see Table 10) to accurately produce simple past tense verbs 80% of the time on the pretest (Essay 1). By the time of the posttest (Essay 3), they are predicted to accurately use the simple past tense 88% of the time. In other words, they achieve close to a 50% reduction in error, as was seen in the direct feedback group. Yet, at the time of the delayed posttest, the metalinguistic feedback group was only predicted to accurately use the simple past tense 81% of the time. Thus, the significant gains seen immediately following the treatment were not maintained at the time of the delayed posttest (Essay 4).

The control group, however, exhibited no significant changes in accuracy between the time of the pretest (78%), the posttest (79%, $\beta = .08$, $SE = 0.21$, $p = .708$), and the delayed posttest (78%, $\beta = -0.12$, $SE = 0.19$, $p = .546$). To recap, this group shows essentially no change in accuracy over time. They are predicted to accurately produce the simple past tense 78% of the time at the time of the pretest (Essay 1), 79% of the time at the time of the posttest (Essay 3), and 78% of the time at the time of the delayed posttest (see Table 10). This lack of change in predicted error rate helps to confirm that the changes seen over time in the two feedback groups are indeed the result of treatment.

To summarize, both the direct and metalinguistic groups exhibited a significant improvement in accuracy from the time of the pretest to the time of the posttest; however, the metalinguistic group failed to maintain these gains, whereas the direct feedback group was able to maintain improvement on a new writing task four weeks later. The control group showed no significant changes in accuracy over time.

Table 10

Predicted Probabilities (log odds with percentage correct in parentheses) of Accurate use of Simple Past Tense by Group at Each Essay Time

	Pretest (Essay 1)	Posttest (Essay 3)	Delayed Posttest (Essay 4)
Direct Feedback Group	.78 (69%)	1.65 (84%)	1.54 (82%)
Metalinguistic Feedback Group	1.37 (80%)	2.0 (88%)	1.44 (81%)
Control Group	1.25 (78%)	1.32 (79%)	1.13 (78%)

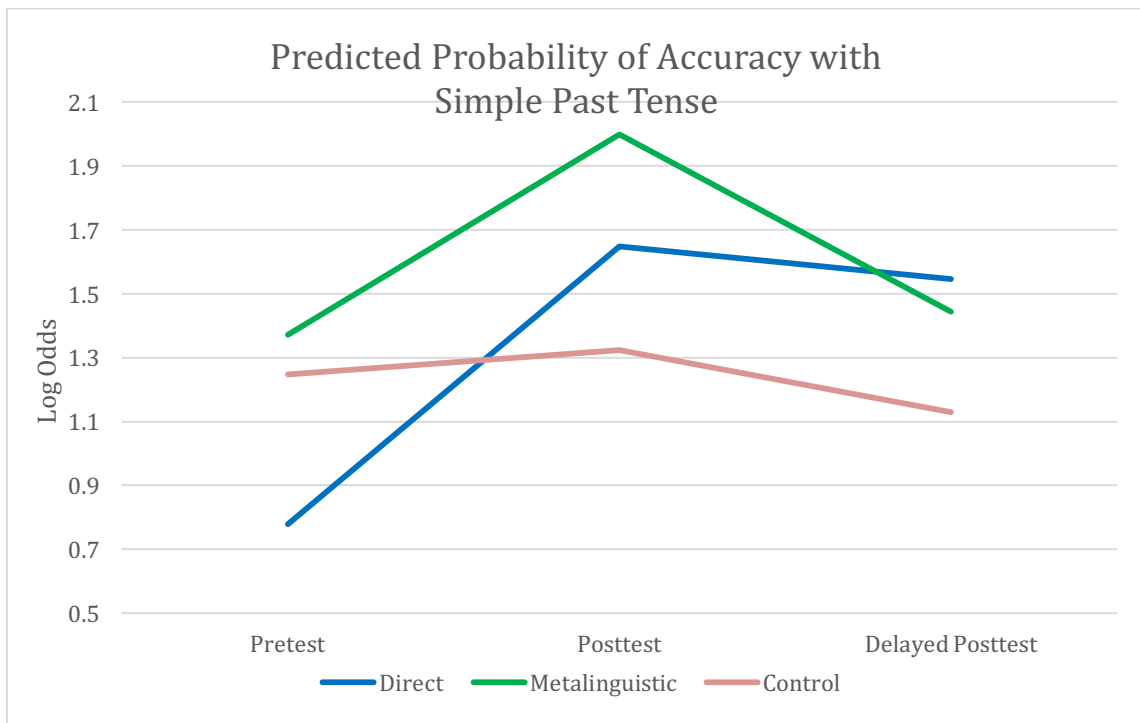


Figure 2. Plot of predicted probabilities (log odds) of accurate use of simple past tense by group at each essay time.

For the direct group, there was also a significant two-way interaction between LLAMA F and Time, in particular for the change in trial-level accuracy from Time 1 to Time 3 ($\beta = .42, SE = .20, p = .035$). Likewise, for the metalinguistic group, there was a

significant two-way interaction between LLAMA F and Time, in particular for gains at Time 3 ($\beta = -0.60$, $SE = .23$, $p = .009$). However, the two-way interaction for the control group between LLAMA F and Time (at Time 3) was not significant ($\beta = -0.18$, $SE = .21$, $p = .387$).

Significant three-way interactions between essay time, LLAMA, and group confirm that the effect of LLAMA seen by the direct group at Time 3 is significantly different than the effect of LLAMA seen by the metalinguistic group ($\beta = -1.02$, $SE = .31$, $p = .001$), and that the effect of LLAMA for the direct group at Time 3 is significantly different than the effect of LLAMA seen by the control group ($\beta = -.60$, $SE = .29$, $p = .038$).

However, there was no three-way interaction between Time, LLAMA, and metalinguistic vs. control groups ($\beta = .42$, $SE = .31$, $p = .179$) at Time 3, meaning that the effect of LLAMA seen in the metalinguistic group may not be different from what is seen in the control group.

There were no significant interactions between LLAMA, group, and the differences at Time 4 (delayed posttest).

These results suggest a possible Aptitude-Treatment Interaction (ATI). A participant in the direct feedback group who is one standard deviation above the mean is predicted by this model to see an additional increase in the probability of correct usage of .42 at the time of the posttest, which adds up to a total probability of 2.07 or 89% accuracy compared to someone with a mean LLAMA F score, who is predicted to achieve 84% accuracy. However, a participant in the metalinguistic group who is one standard deviation above the mean LLAMA F score on the pretest is predicted to see a

decrease in probability correct of $-.60$ on the posttest, lowering him or her to a predicted probability correct of 1.39 or 80% accuracy, which is lower than what is predicted at the time of the posttest for someone with a mean LLAMA F score, who is predicted to achieve 88% accuracy. In other words, learners in the direct group begin with 69% accuracy. If a learner in this group has a LLAMA F score that is one standard deviation above the mean, he or she would be expected to achieve 89% accuracy at Time 3 (a 5% boost from someone with an average LAA), whereas with metalinguistic feedback a learner with a LLAMA F score one standard deviation above the mean would begin with 80% accuracy and remain at 80% accuracy after two rounds of feedback, essentially seeing no improvement. In summary, for the direct feedback group, as LLAMA F score increases, accuracy increases, yet for the metalinguistic group, accuracy decreases as LLAMA F score increases (see Figure 3). At the time of the delayed posttest, there was still an increase in predicted accuracy of 0.10 for participants with higher LLAMA F scores in the direct feedback group, but this benefit of LLAMA F was no longer significant. The decrease in accuracy seen by the metalinguistic group at the time of the delayed posttest for those with a LLAMA F score one standard deviation above the mean was less at -0.16 and was also no longer significant.

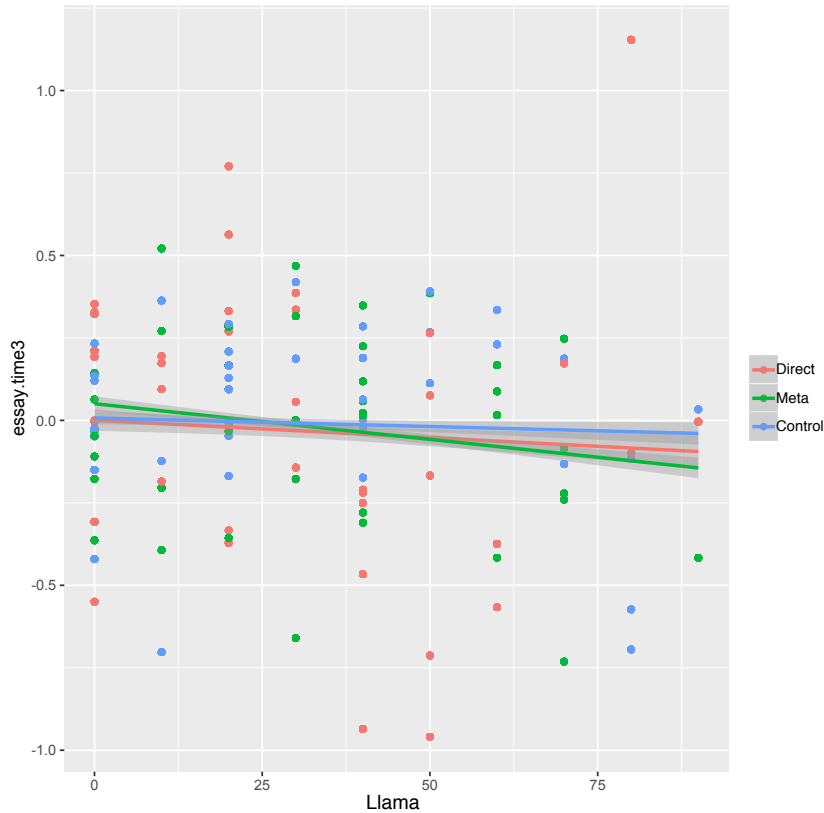


Figure 3. Effect of LLAMA F at Essay Time 3.

In order to test the differential effect of treatment depending on aptitude, additional models were fit with standardized LLAMA F scores that were shifted two standard deviations above and below the mean to create new variables to allow for comparisons between the effects of these two types of feedback on learners with higher and lower levels of language aptitude.

Learners with higher LLAMA F scores, specifically two standard deviations above the mean, in the direct group began with 57% predicted accuracy, and learners in the metalinguistic group began with 88% accuracy. This difference in accuracy between groups at the start of the study was significant ($\beta = 1.73, SE = .61, p = .005$). Learners with higher LLAMA F scores in the direct group significantly increased their rate of

predicted accuracy at Time 3 (88%, $\beta = 1.708$, $SE = .45$, $p = .000$), and although somewhat lower, these learners' gains were still significant at Time 4 (78%, $\beta = .973$, $SE = .41$, $p = .018$). In contrast, learners with higher LLAMA F scores in the metalinguistic group began with 88% accuracy, yet these learners are predicted by the model to decrease in accuracy to 81% at Time 3 ($\beta = -0.580$, $SE = .51$, $p = .254$) and to 85% at Time 4 ($\beta = -0.252$, $SE = .46$, $p = .584$). Neither decrease, however, was significant.

However, for learners with higher LLAMA F scores, the differences in gains between the direct and metalinguistic groups at Time 3 ($\beta = -2.288$, $SE = .68$, $p = .001$) and Time 4 ($\beta = -1.225$, $SE = .62$, $p = .048$) was significant. These results suggest that learners with a higher LAA would benefit more from direct feedback than from metalinguistic feedback.

On the other hand, learners with lower LLAMA F scores (two standard deviations below the mean) were predicted by the model to begin with 78% accuracy in the direct group and 68% accuracy in the metalinguistic group. This initial difference in accuracy was not significant ($\beta = -0.543$, $SE = .61$, $p = .374$). Learners with lower LLAMA scores in the direct group were predicted to make slight gains in accuracy from Time 1 (78%) to Time 3 (79%, $\beta = .031$, $SE = 0.45$, $p = .945$). These gains were not significant. This group increased again slightly at Time 4, but the gain was still not significant (86%, $\beta = .558$, $SE = 0.419$, $p = .182$). In contrast, learners with lower LLAMA scores in the metalinguistic group were predicted to benefit from metalinguistic feedback. They significantly improved from Time 1 (68%) to Time 3 (93%, $\beta = 1.83$, $SE = .52$, $p = .000$); however, their accuracy rate decreased to 76% at Time 4 ($\beta = .393$, $SE = .45$, $p = .380$). The difference in accuracy rate between Time 1 and Time 4 was not significant.

The difference in gains at Time 3 between the metalinguistic and direct feedback groups was significant ($\beta = 1.799$, $SE = .68$, $p = .009$). However, the difference in gains between these two groups at Time 4 was not significant ($\beta = -0.165$, $SE = .61$, $p = .787$).

Thus, at the time of the immediate written posttest (Time 3), this model suggests that learners with lower LAA benefitted more from metalinguistic feedback than direct feedback. However, this effect was no longer present at the time of the delayed written posttest (Time 4).

In summary, these comparisons imply that learners with lower LAA will be more successful with metalinguistic feedback, whereas learners with higher LAA will benefit more from direct feedback.

5.3.1 Regular vs. irregular verbs

In order to determine whether or not there was a difference in accuracy gains for regular verbs and irregular verbs, an additional model was fit which included verb type (regular or irregular) and the interaction between verb type and essay time in addition to all of the terms in the best model reported above (pretest scores, LLAMA F scores, group, essay time, and the interaction between group and essay time, as well as interactions between LLAMA and group, LLAMA and essay time, and LLAMA, group, and essay time as fixed effects, and random effects of participant, including the random intercept and random slopes for the effects of essay time).

There were 2292 total trials for regular verbs compared to 5814 total trials for irregular verbs. Table 11 offers a breakdown of the number of trials for both regular and irregular verbs for each group and the percentage correct for each verb type.

Table 11

Total Number of Trials and Percentage Correct for Regular and Irregular Simple Past Tense

Group	regular verb trials	irregular verb trials
Direct	853 (67%)	2121 (78%)
Metalinguistic	736 (74%)	1895 (84%)
Control	703 (62%)	1798 (78%)

The difference in irregular verb error rates compared to regular verb error rates was significant for the direct group ($\beta = 0.87$, $SE = .22$, $p < .001$) and the control group ($\beta = 1.13$, $SE = .23$, $p < .001$), and marginally significant for the metalinguistic group ($\beta = 0.48$, $SE = .26$, $p = 0.06$). However, there was no significant interaction found between verb type and essay time for any of the groups ($p > .05$), which indicates that the gains made for regular verbs as compared to the gains made for irregular verbs were not significantly higher for any of the groups at Essay Time 3 or Time 4. This model does tell us that the probability of accuracy for irregular verbs for the direct feedback group continued to be slightly lower at Essay Time 3 ($\beta = -0.19$, $SE = .30$, $p = .516$) and Time 4 ($\beta = -0.52$, $SE = .31$, $p = 0.09$) than the probability of accuracy for regular verbs. Similarly, for the control group, the probability for accuracy for irregular verbs was slightly lower at Essay Time 3 ($\beta = -0.38$, $SE = .32$, $p = 0.23$) and Time 4 ($\beta = -0.50$, $SE = .32$, $p = .120$) than the probability of accuracy with regular verbs.

Conversely, the probability of accuracy for irregular verbs for the metalinguistic feedback group was slightly higher at Essay Time 3 ($\beta = 0.34$, $SE = .35$, $p = .329$) and

Time 4 ($\beta = 0.50$, $SE = .34$, $p = .142$) than the probability of accuracy with regular verbs. However, none of these differences in probability were significant.

5.4 Present Perfect Tense in the Writing Tasks

All 151 participants made at least a few errors with the present perfect tense. However, seven participants deviated from the writing prompts and were able to avoid using the perfect tense on two or more essays. These participants were deleted from the data set since they did not receive two rounds of feedback on the target structures. A total of 2,480 observations from 144 participants were included in the analysis for the present perfect tense. A breakdown of the number of observations at each essay time for each group is presented in Table 12. Percentage correct for the present perfect tense for each essay and each group are given in Table 13, and the number of correct and incorrect uses of present perfect tense are visually presented in Figure 4.

Table 12

Total Number of Obligatory Contexts for Present Perfect at Each Essay Time

	Essay 1	Essay 2	Essay 3	Essay 4	Total
Direct feedback group ($n = 47$)	190	231	247	192	860
Metalinguistic feedback group ($n = 45$)	168	218	240	211	837
Control group ($n = 47$)	195	209	173	206	783
Totals for all groups	553	658	660	609	2480

Table 13

Percentage Correct for Present Perfect for Each Essay Time for Each Group

	Essay 1 % correct	Essay 2 % correct	Essay 3 % correct	Essay 4 % correct
Direct	37	32	55	40
Meta	42	35	56	42
Control	47	41	36	36

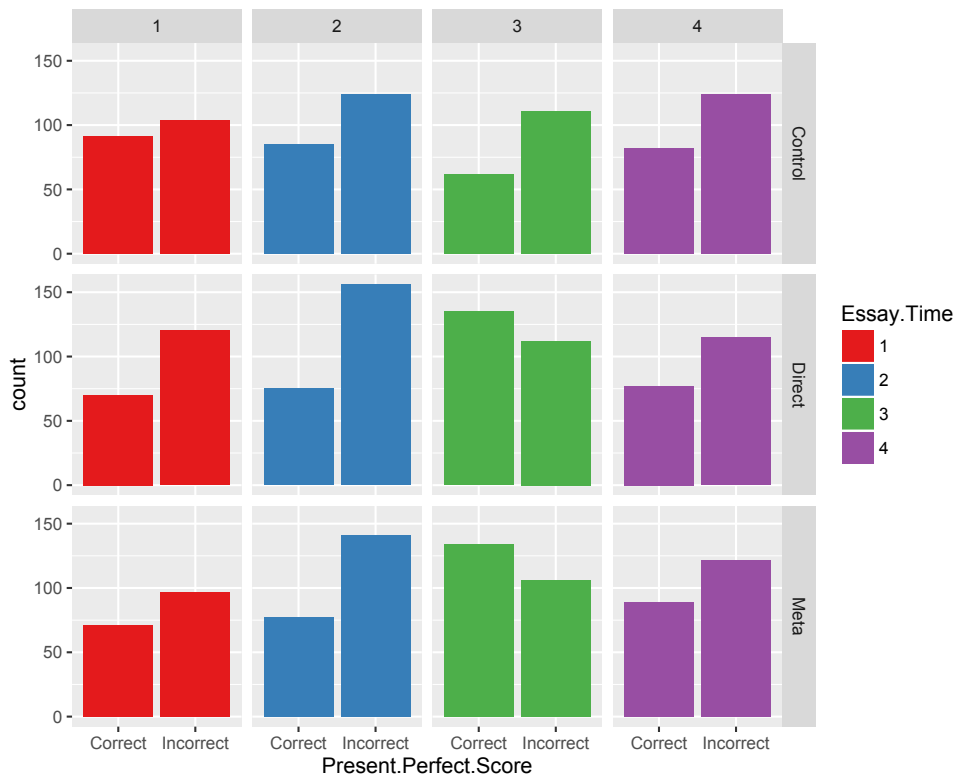


Figure 4. Correct and incorrect trials with present perfect tense for each group at each essay time.

For the present perfect data analysis, the outcome variable was accurate use of the present perfect tense. For each essay, the number of obligatory instances of present perfect were coded with a 1 when correct and a 0 when incorrect. Verbs were considered incorrect both when the wrong tense was chosen (e.g., simple past was used in lieu of present perfect), and when there was an error with form (e.g., *have caught* instead of *have caught*). As with the simple past tense, the predictors in the model included group membership (direct, metalinguistic, or control), centered LLAMA F scores, and centered scores from the verb use section of the form-focused grammar pretest. LLAMA F scores

for the participants included in this analysis ranged from 0 to 90 with a mean of 32.2 and a median of 30. Grammar scores for the verb section of the pretest ranged from 3 to 10 with a mean of 7.877 and a median of 8.

To determine whether or not there was a main effect for group on accurate use of the present perfect tense, a mixed-effects logit model with group, essay time, and the interaction between group and essay time as fixed effects and subject as a random effect was fit. The first model, which included only random intercept, was compared to one that included both random intercept and random slope. As was the case with the simple past tense, the results of a comparison between models showed that random slopes significantly improved model fit, $\chi^2(6) = 29.61, p < .001$.

Additional models, which included standardized LLAMA F scores, standardized form-focused pretest scores, and interactions with these effects, were fit and compared. For this structure, no significant interaction with LLAMA F score was found, and LLAMA F score alone did not improve model fit. However, the form-focused pretest scores once again accounted for a significant amount of individual variance and were included. The best model, reported below and presented in Table 14, includes pretest scores, group, essay time, the interaction between group and essay time as fixed effects. The model was also fit with random effects of participant, including the random intercept and random slopes for the effects of essay time.

Table 14

Result Summary of Coefficient Estimates β , Standard Errors SE, Associated Wald's z-score, and Significance Level p for Fixed and Random Effects in the Present Perfect Tense Analysis

	Coef. β	SE	Z	P	Participant (SD)
Intercept (Control)	-0.32	0.25	-1.33	0.183	0.87
Pretest	0.36	0.11	3.37	0.001 ***	
Direct	-0.51	0.35	-1.45	0.147	
Metalinguistic	-0.31	0.35	-0.89	0.375	
Essay Time 3	0.50	0.32	-1.56	0.118	0.68
Essay Time 4	-0.30	0.32	-0.95	0.342	1.32
Essay Time 3:Direct	1.38	0.44	3.17	0.002 **	
Essay Time 3:Meta	1.33	0.44	3.05	0.002 **	
Essay Time 4:Direct	0.20	0.45	0.45	0.650	
Essay Time 4:Meta	0.34	0.44	0.78	0.438	

Note: * = $p < .05$; ** = $p < .01$; *** = $p < .001$

Although slightly lower, the direct feedback group's predicted accuracy (30%) at Essay Time 1 was not significantly less than the metalinguistic group's accuracy (35%, $\beta = .20$, $SE = 0.36$, $p = .579$), nor was this group significantly lower than the control group (42%, $\beta = .51$, $SE = .35$, $p = .147$). The metalinguistic and control groups were likewise not significantly different from each other at Time 1 ($p > .05$). In short, there were no significant differences between groups at the start of the study.

Learners who received direct corrective feedback made significant gains from the time of the pretest (30%) to the time of the immediate posttest (51%, $\beta = .88$, $SE = 0.31$, $p = .004$). These gains, however, were not maintained at the time of the delayed posttest (28%, $\beta = -0.09$, $SE = 0.34$, $p = .779$). As shown in Table 15, the direct feedback group is

predicted to accurately produce present perfect tense 30% of the time on the pretest (Essay 1). By the time of the posttest (Essay 3), these learners are predicted to accurately use the present perfect tense 51% of the time. At the time of the delayed posttest, however, the direct feedback group is predicted to accurately supply the present perfect tense only 28% of the time. Thus, the significant gains that were achieved at the time of the immediate posttest were lost by the time of the delayed posttest (Essay 4).

The metalinguistic feedback group also made significant gains in accuracy from the time of the pretest (35%) to the time of the immediate posttest (55%, $\beta = .83$, $SE = 0.30$, $p = .006$). However, as was the case with the direct feedback group, learners who received metalinguistic feedback were not able to maintain these gains by the time of the delayed posttest (36%, $\beta = 0.05$, $SE = 0.32$, $p = .888$). As shown in Table 15, the metalinguistic group is predicted to produce accurate present perfect tense 35% of the time on the pretest (Essay 1). By the time of the posttest (Essay 3), they are predicted to accurately supply the present perfect tense 55% of the time. At the time of the delayed posttest, the metalinguistic feedback group reverts back to accurate use the present perfect tense only 36% of the time meaning that the significant gains that were predicted at the time of the immediate posttest were lost at the time of the delayed posttest (Essay 4).

This model further confirms that the gains seen at Time 3 by the direct group are significant in comparison to the control group ($\beta = 1.39$, $SE = .44$, $p = .002$), and that likewise the gains achieved by the metalinguistic group at Time 3 were significant in comparison to the control group ($\beta = 1.33$, $SE = .44$, $p = .002$).

The control group showed no significant changes in accuracy between the time of the pretest (42%), the posttest (31%, $\beta = -.50$, $SE = 0.32$, $p = .118$), and the delayed posttest (35%, $\beta = -.30$, $SE = 0.32$, $p = .342$). This group is predicted to accurately produce the present perfect tense 42% of the time at the time of the pretest (Essay 1), yet these participants drop to 31% accuracy at the time of the posttest (Essay 3), and then they rebound slightly to 35% at the time of the delayed posttest (see Table 15).

To summarize, both the direct and metalinguistic groups saw a significant improvement in accuracy with the present perfect tense from the time of the pretest to the time of the posttest; however, neither treatment group was able to maintain these gains four weeks later on the delayed written posttest. The control group showed no significant changes in accuracy at any point.

Table 15

Predicted Probabilities (log odds with percentage correct in parentheses) of Accurate use of the Present Perfect Tense

	Pretest (Essay 1)	Posttest (Essay 3)	Delayed Posttest (Essay 4)
Direct Feedback Group	-.84 (30%)	.05 (51%)	-.93 (28%)
Metalinguistic Feedback Group	-.64 (35%)	.19 (55%)	-.59 (36%)
Control Group	-.33 (42%)	-.82 (31%)	-.63 (35%)

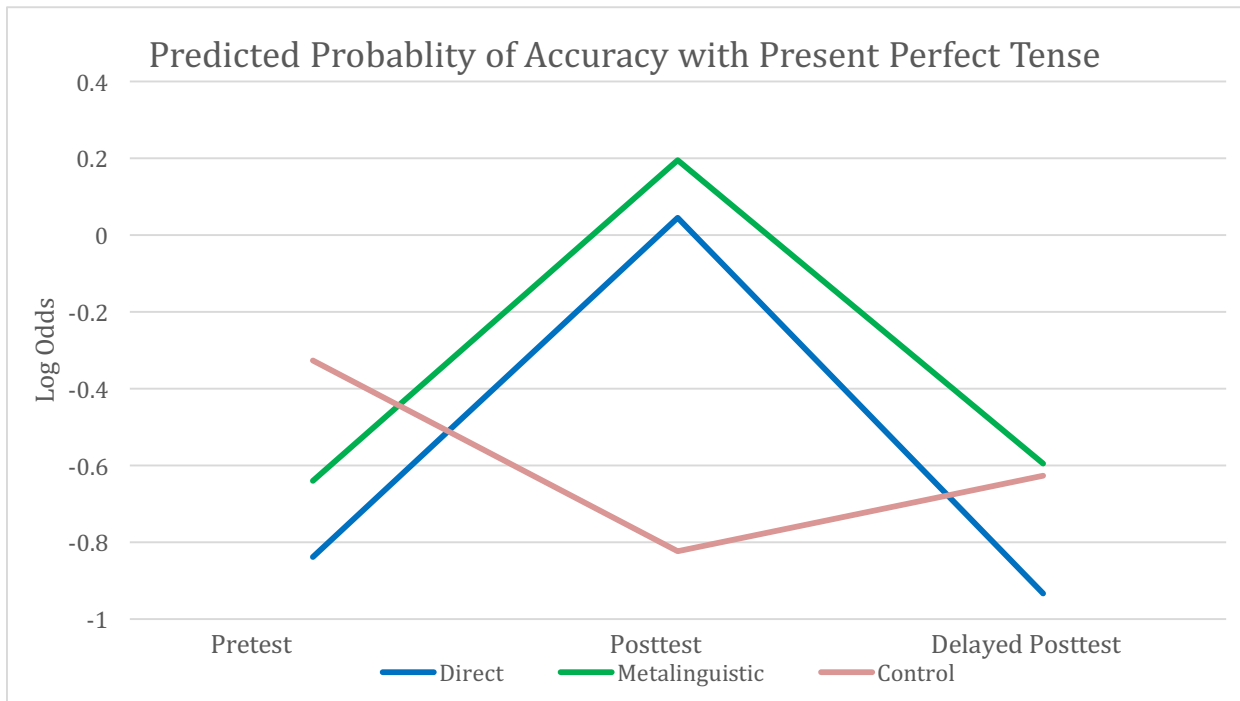


Figure 5. Plot of predicted probabilities of accurate use of present perfect tense by group at each essay time.

5.5 Dropped Pronouns in the Writing Tasks

The results obtained from the analysis for dropped pronouns should be interpreted with caution. Only 50 participants consistently dropped pronouns during the study. Consistent was defined as three or more instances of dropped pronouns within Essay 1 and Essay 2 combined. Participants who dropped less than three pronouns were excluded from the analysis since these were likely random mistakes rather than linguistic error. The largest group of pronoun droppers were Spanish speakers ($n = 28$). There were also five Arabic, six Vietnamese, two Portuguese, one Albanian, one Bulgarian, one Chinese, one Croatian, one Punjabi, one Russian, one Tagalog, one Thai, and one Ukrainian speaker who consistently omitted subjects. This small subset of participants was examined to determine whether or not written corrective feedback reduced the number of dropped pronouns over time.

First, the form-focused grammar test scores for this subset of participants were analyzed. The descriptive statistics are presented in Table 16. A mixed-effects ANOVA with treatment group as a between-subjects factor and essay time as a within-subjects factor was conducted to compare group means on the scores for the pronoun dropping section of the form-focused pretest and the posttest for this subset. There was no significant effect for group ($F(2, 47) = .384; p > .05$).

Table 16

Descriptive Statistics for the Dropped Pronoun Section of the Form-focused Grammar Test (SD in parentheses) for Data Subset

Group	N	Mean	SD
Direct	14		
Pretest		8.00	1.66
Posttest		8.14	2.11
Metalinguistic	21		
Pretest		7.91	2.39
Posttest		8.24	1.81
Control	15		
Pretest		7.47	2.39
Posttest		7.07	2.96

Although the scores for the direct feedback group ($F(1, 13) = .125 p > .05$) and the scores for the metalinguistic group ($F(1, 20) = .352 p > .05$) increased slightly from pretest to posttest, these gains were not significant. The scores for the control group ($F(1, 14) = .251 p > .05$) decreased slightly from pretest to posttest, but likewise, this change was not significant.

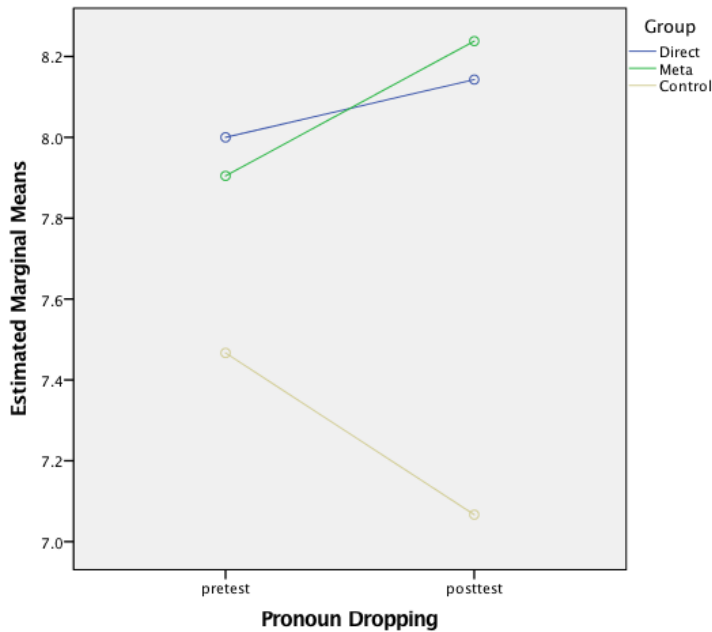


Figure 6. Estimated marginal means from pretest to posttest for pronoun dropping section for subset of learners.

6,997 observations from 50 participants were included in the mixed-effects logistic regression analysis for dropped pronouns. Table 17 presents a breakdown of the number of observations and the percentage of correctly supplied subjects by group at each essay time.

Table 17

Number of Observations (Obligatory Use of a Subject) at Each Essay Time and Percentage Correct

	Essay 1 Subjects	Essay 2 Subjects	Essay 3 Subjects	Essay 4 Subjects
Direct feedback group (n = 14)	516 (95%)	520 (94%)	532 (96%)	536 (98%)
Metalinguistic feedback group (n = 21)	749 (93%)	735 (96%)	746 (97%)	649 (96%)
Control group (n = 15)	565 (93%)	554 (94%)	456 (92%)	439 (96%)

For this analysis, the outcome variable was subject provision. For each essay, the number of obligatory subjects was coded with a 1 when a subject or pronoun was provided, and a 0 when the subject or pronoun was missing. The predictors for the model were group membership (direct, metalinguistic, or control), the total word count for each essay, LLAMA F scores, and scores from the pronoun dropping section of the form-focused grammar pretest. LLAMA scores for this data subset ranged from 0 to 80 with a mean of 26.86 and a median of 20. As before, LLAMA F scores and the grammar pretest scores were standardized. Total word count per essay ranged from 99 to 824 with a mean of 429 and a median of 407. Word count was likewise standardized.

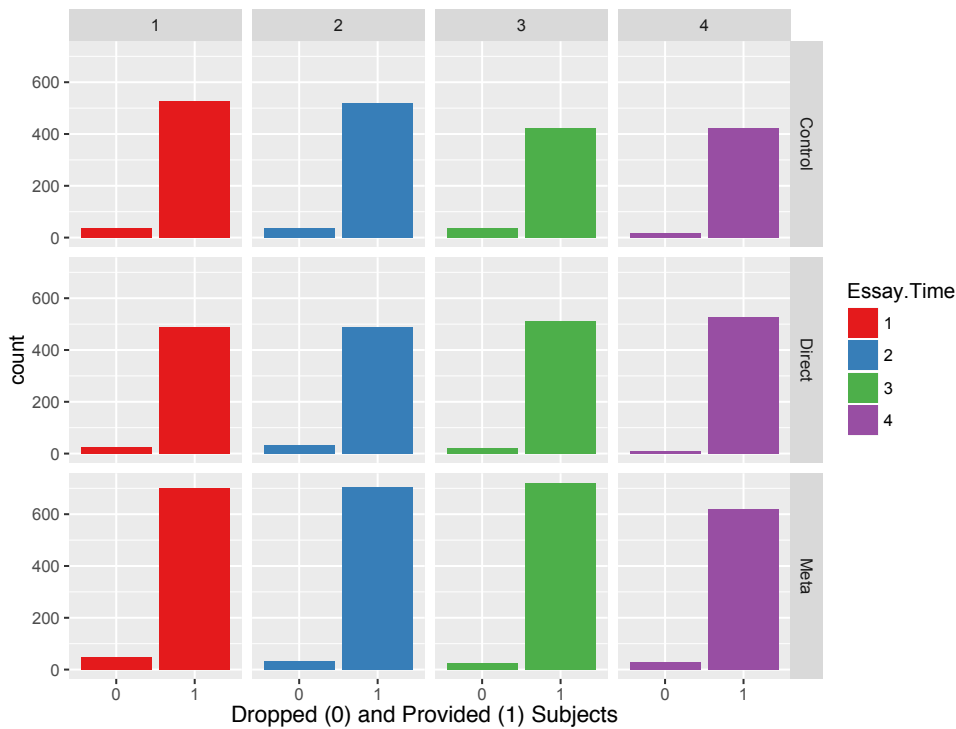


Figure 7. Number of missing pronouns (0s) and accurately supplied pronouns (1s) for each group at each essay time.

To examine the effect of group on accurate provision of obligatory subjects or pronouns, a mixed-effects logit model with group, essay time, and the interaction between group and essay time as fixed effects was fit. The model was further fit with random effects of participant, including the random intercept and random slopes for the effects of essay time. As was the case with the simple past tense and the present perfect tenses analyses, a model, which included only random intercept, was compared to one that included both random intercept and random slope. Not surprisingly, the results of a comparison between models, showed that random slopes significantly improved model fit, $\chi^2(10) = 28.46, p = 0.002$.

Additional models, which included standardized LLAMA F scores, standardized form-focused pretest scores, and interactions with these effects were fit and compared. For this structure, no significant interaction with LLAMA F score was found, and the inclusion of LLAMA F scores did not improve model fit. However, the form-focused pretest scores significantly improved model fit and were included. Standardized word count for each essay was also added to the model and was found to be a significant predictor. The best model, reported below, includes pretest scores, word count, group, essay time, the interaction between group and essay time as fixed effects, and random effects for participant, including the random intercept and random slopes for essay time.

Table 18

Result Summary of Coefficient Estimates β , Standard Errors SE, Associated Wald's z-score, and Significance Level p for Fixed and Random Effects in the Dropped Pronoun Analysis

	Coef. β	SE	z	P	Participant (SD)
Intercept (Control)	2.72	0.19	14.45	0.000 ***	0.00
Pretest	0.20	0.07	3.06	0.002 **	
Word Count	0.29	0.08	3.43	0.001 ***	
Direct	0.24	0.28	0.85	0.395	
Metalinguistic	0.10	0.25	0.40	0.693	
Essay Time 2	0.16	0.36	0.43	0.666	0.75
Essay Time 3	0.03	0.33	0.11	0.917	0.69
Essay Time 4	1.10	0.43	2.56	0.010 *	0.87
Essay Time 2:Direct	-0.31	0.52	-0.58	0.560	
Essay Time 2:Meta	0.47	0.49	0.96	0.336	
Essay Time 3:Direct	0.34	0.49	0.71	0.480	
Essay Time 3:Meta	0.81	0.44	1.82	0.069 .	
Essay Time 4:Direct	0.29	0.64	0.45	0.649	
Essay Time 4:Meta	-0.30	0.53	-0.56	0.578	

Note: * = $p < .05$; ** = $p < .01$; *** = $p < .001$

At the start of the study, the control group (94%) was not significantly different from the direct group (95%, $\beta = .24$, $SE = 0.28$, $p = .395$), nor was this group significantly different from the metalinguistic group ($\beta = .10$, $SE = 0.25$, $p = .693$). The metalinguistic group and direct group were likewise not significantly different ($p > .05$). In sum, there were no significant differences between groups at the start of the study.

Learners who received direct feedback improved slightly from the time of the written pretest (95%) to the posttest (97%, $\beta = .38$, $SE = 0.38$, $p = .322$); however, these

gains were not significant. Yet, by the time of the delayed written posttest, the direct feedback group had achieved a significant boost in probable accuracy (99%, $\beta=1.39$, $SE = 0.51$, $p = .006$.) As shown in Table 19, the error rate is low to begin with for all groups. The direct feedback group begins with 95% accuracy at essay time one and 94% accuracy at essay time two. By the time of the immediate posttest (Essay 3), this group sees some improvement in predicted accuracy with an accuracy rate of 97%. This improvement, however, is not significantly different from the accuracy rate predicted at the start of the study. Yet, by the delayed posttest (Essay 4), the direct group is predicted to accurately supply subjects 99% of the time, and this improvement was significant.

In comparison, the metalinguistic feedback group improved significantly from the time of the pretest (94%) to the posttest (97%, $\beta = 0.84$, $SE = 0.33$, $p = .010$). These gains were maintained at the time of the delayed posttest (97%, $\beta=0.80$, $SE = 0.37$, $p = .030$). The metalinguistic feedback group begins with 94% accuracy at essay time one, which increases slightly to 97% accuracy at essay time two, after one round of feedback. After receiving two rounds of feedback, at the time of the immediate posttest (Essay 3), learners who received metalinguistic feedback see a significant improvement (compared to time one) in predicted accuracy (97%). As reported in Table 19, this gain in accuracy is maintained at the time of the delayed posttest (Essay 4).

Learners who did not receive feedback did not significantly increase the probability of a correct outcome from the time of the pretest (94%) to the posttest ($\beta = 0.035$, $SE = 0.33$, $p = .917$). Yet, by the time of the delayed posttest, the control group did achieve a significant boost in probability of a correct outcome (98%, $\beta=1.10$, $SE = 0.43$, $p = 0.010$). Like the direct feedback group, this group sees no significant change in

accuracy rate from Essay 1 at 94%, to Essay 2 at 95%, to Essay 3 at 94%; however, interestingly, at Essay time 4, this group makes a significant gain in predicted accuracy (98%).

In comparison to the control group, only the gains seen at Time 3 by the metalinguistic group are statistically different; the change in accuracy achieved by learners who received metalinguistic feedback is only marginally different from what occurs in the control group ($\beta = 0.81, SE = .44, p = .068$).

Due to the small number of participants who made this error and the small numbers of errors made by most of the participants, these results should be considered exploratory. It appears that there is some improvement in accuracy after treatment for the direct and metalinguistic groups; the gains at Essay Time 3 for the metalinguistic group are significant; however, the fact that all three groups, including the control group, are able to significantly improve by the end of the semester suggests that something other than treatment may be involved.

Table 19

Predicted Probabilities (log odds with percentage correct in parentheses) of a Correctly Supplied Pronoun

	Pretest (Essay 1)	Essay 2 (after feedback on Essay 1)	Posttest (Essay 3)	Delayed Posttest (Essay 4)
Direct Feedback Group	2.96 (95%)	2.81 (94%)	3.34 (97%)	4.35 (99%)
Metalinguistic Feedback Group	2.82 (94%)	3.44 (97%)	3.66 (97%)	3.62 (97%)
Control Group	2.72 (94%)	2.88 (95%)	2.75 (94%)	3.82 (98%)

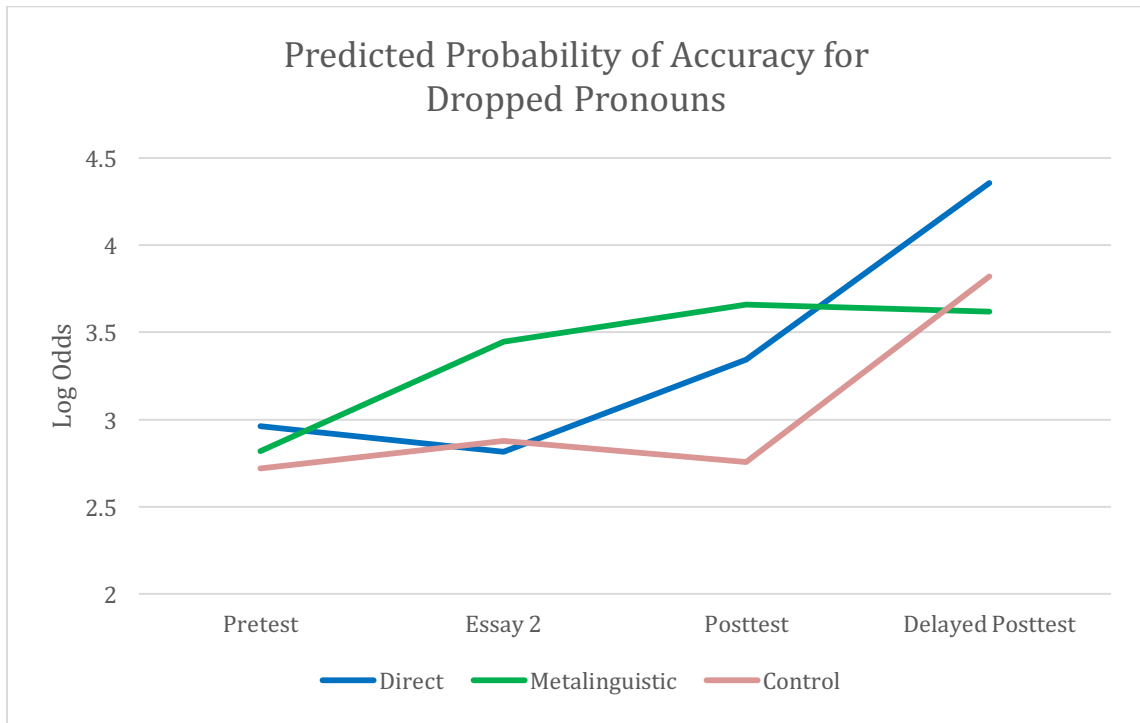


Figure 8. Plot of predicted probabilities of accuracy for dropped pronouns by group at each essay time.

5.6 Pronominal Duplication in the Writing Tasks

Very few participants consistently included redundant pronouns in their writing. Only those who included two or more redundant pronouns throughout the course of the study and received feedback at least once, at either Essay Time 1 or Essay Time 2 were included in the analysis ($n = 21$). The results of this analysis, therefore, should be interpreted with considerable caution as the number of observations is very small. The largest L1 groups for pronominal redundancy were Arabic speakers ($n = 8$) and Spanish speakers ($n = 8$). There was also one Catalan speaker, two Greek speakers, one Kyrgyz speaker, and one Thai speaker who consistently supplied redundant pronouns.

The analysis for pronominal duplication was different from those reported thus far, since the outcome variable did not have a binomial distribution. The outcome

variable was the total count in each essay of redundant pronouns; therefore, a Poisson model, which is appropriate for count data, was used.

The potential predictors in the model included group membership (direct, metalinguistic, or control), the total number of required subjects for each essay, word count for each essay, LLAMA F scores, and scores from the pronominal duplication section of the form-focused grammar pretest. LLAMA F scores, grammar pretest scores, word count, and number of required subjects were standardized. LLAMA F scores for this small subset ranged from 0 to 80, with a mean of 21.90 ($SD = 24.21$), and a median of 20. The total number of obligatory subject use per text ranged from 10 to 93 with a mean of 39.96 and a median of 36.50. The count for redundant pronoun use ranged from 0 to 10 with a mean of 1.452 and a median of 1. Total word count per essay ranged from 99 to 824 with a mean of 410.8 and a median of 394.

Table 20

Descriptive Statistics for Variables Used in Pronominal Duplication Analysis

	Min	Max	Mean	SD
Redundant Count	0	10	1.45	1.609
Required Subjects	10	93	39.96	18.024
Word Count	99	824	410.81	163.136
LLAMA F scores	0	80	21.90	23.770

The form-focused grammar test scores for this subset of participants were analyzed. The descriptive statistics are presented in Table 21. A mixed-effects ANOVA with treatment group as a between-subjects factor and essay time as a within-subjects factor was conducted to compare group means on the scores for the pronoun duplication section

of the form-focused pretest and the posttest for this subset. There was no significant effect for group ($F(2, 18) = 1.053; p > .05$).

Table 21

Descriptive Statistics for Pronominal Duplication Section of Form-focused Grammar Test (SD in parentheses) for Data Subset

<u>Group</u>	<u>N</u>	<u>Mean</u>	<u>SD</u>
Direct	7		
Pretest		6.29	2.81
Posttest		6.00	2.08
Metalinguistic	9		
Pretest		7.22	2.68
Posttest		8.00	1.94
Control	5		
Pretest		6.80	3.27
Posttest		7.60	2.88

Although the scores for the metalinguistic feedback group ($F(1, 8) = 2.227 p > .05$) and the scores for the control group ($F(1, 4) = 4.571 p > .05$) increased slightly from pretest to posttest, these gains were not significant. The scores for the direct feedback group ($F(1, 6) = .146 p > .05$) decreased slightly from pretest to posttest, but likewise, this change was not significant.

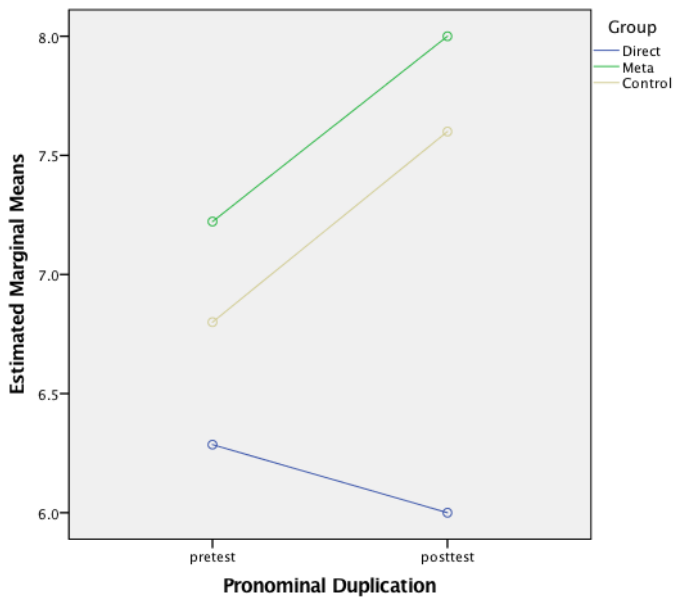


Figure 9. Estimated marginal means for pretest and posttest scores for pronominal duplication section of form-focused test for subset of learners.

To examine the effect of group on accurate provision of obligatory subjects or pronouns, a mixed-effects model was fit with group, essay time, and the interaction between group and essay time as fixed effects and the random effects of participant. The first model, which only included random intercept, was compared to one that included both random intercept and random slopes for the effects of essay time. In contrast to the models fit for the other structures, the results of a comparison between models showed that the addition of random slopes did not significantly improve model fit, $\chi^2(10) = 8.31$, $p = 0.60$, and therefore, the final model reported below includes only random intercept. One by one the other predictors (standardized LLAMA F scores, standardized form-focused pretest scores, standardized word count, and total number of subjects) were included in the model, and likelihood ratio tests were conducted after each addition to check for improvements in model fit. Due to the extremely small number of observations for this structure, it was not possible to test for an interaction with LLAMA F score, and

the inclusion of LLAMA F scores on their own did not improve model fit. However, the form-focused pretest scores significantly improved model fit and were included.

Standardized subject count for each essay was also found to be a significant predictor.

Word count was strongly correlated with the subject count and therefore did not improve model fit when included. The best model, reported below, includes pretest scores, number of subjects, group, essay time, and the interaction between group and essay time as fixed effects, and random effects for participant (random intercept only).

Table 22

Result Summary of Coefficient Estimates β , Standard Errors SE, Associated Wald's z-score, and Significance Level p for Fixed and Random Effects in the Pronominal Duplication Analysis

	Coef. β	SE	z	p	Participant (SD)
Intercept (Control)	-0.59	0.49	-1.20	0.230	0.18
Pretest.scaled	-0.27	0.10	-2.63	0.009 **	
Subjects.scaled	0.02	0.01	2.90	0.004 **	
Direct	0.43	0.54	0.79	0.433	
Metalinguistic	0.87	0.50	1.71	0.087 .	
Essay Time 2	0.25	0.61	0.42	0.676	
Essay Time 3	0.43	0.59	0.73	0.466	
Essay Time 4	0.03	0.63	0.043	0.966	
Essay Time 2:Direct	-1.18	0.78	-1.50	0.134	
Essay Time 2:Meta	-1.63	0.76	-2.15	0.032 *	
Essay Time 3:Direct	-0.93	0.73	-1.27	0.203	
Essay Time 3:Meta	-1.00	0.68	-1.48	0.139	
Essay Time 4:Direct	-0.71	0.79	-0.90	0.368	
Essay Time 4:Meta	-0.30	0.71	-0.42	0.676	

Note: * = p < .05; ** = p < .01; *** = p < .001

At Essay Time 1, the control group (36%) was not significantly different from the direct group (46%, $\beta = 0.43$, $SE = 0.54$, $p = .433$), but the control group was marginally different from the metalinguistic group (57%, $\beta = .87$, $SE = 0.50$, $p = .09$). The metalinguistic group (57%) was not, however, significantly different at essay time one from the direct group (46%, $\beta = -0.44$, $SE = 0.37$, $p = .230$). The control group thus began with the lowest probability of pronominal duplication error (36%), compared to the direct feedback group (46%), and the metalinguistic group (57%), as shown in Table 23.

The direct feedback group improved slightly from the time of the written pretest (46%) to the posttest (34%, $\beta = -0.50$, $SE = 0.43$, $p = .242$); nevertheless, this decrease in error probability was not significant. At the time of the delayed written posttest, the direct feedback group was still trending in the right direction (30%, $\beta = -0.68$, $SE = 0.47$, $p = .146$), but the reduction in errors was still not significant. There was, however, a marginally significant decrease in error probability at Essay Time 2 (25%, $\beta = -0.92$, $SE = 0.50$, $p = .06$), after one round of feedback, in comparison to Essay Time 1. As shown in Table 21, the direct feedback group begins with a probability of including a redundant pronoun error of 46% at the start of the study and drops to only 25% error at essay time two. By the time of the immediate posttest (Essay 3), this group's improvement in predicted error increases back to 34%, which is not significantly different from the probability seen at the beginning. By the time of the delayed posttest (Essay 4), the direct group is at 30% probability, which is still not significantly different from essay time one.

The metalinguistic feedback group improves significantly from the time of the pretest (57%) to Essay Time 2 (25%, $\beta = -1.38$, $SE = 0.46$, $p = .003$). At the time of the immediate posttest (43%, $\beta = -0.57$, $SE = 0.33$, $p = .083$), the reduction in errors is less

than at Essay 2, but still marginally better than at the time of the pretest. At the time of the delayed posttest (50%, $\beta = -0.27$, $SE = 0.33$, $p = .410$), the metalinguistic group's error rate is still lower than at the start of the study, but the decrease is no longer significantly different from what is predicted at Essay 1. In other words, as shown in Table 21, learners who received metalinguistic feedback began with a probability of including a redundant pronoun error of 57% at Essay Time 1 and were able to reduce their error rate to 25% at Essay Time 2. By the time of the immediate posttest (Essay 3), this group's predicted error rate increases back to 43%, which is still marginally better than the probability seen at the beginning of the study. By the time of the delayed posttest (Essay 4), however, the metalinguistic group is at 50% error probability, which is no longer significantly different from time one.

Finally, the control group's rate of error at time one (36%) is the lowest of the three groups. At Essay Time 2 (42%, $\beta = 0.25$, $SE = 0.61$, $p = .676$), this group's error rate probability increases slightly but not significantly. The control sees another slight increase in error probability at time three (46%, $\beta = 0.43$, $SE = 0.59$, $p = .466$), which is still not significantly different from the rate at time one. At the time of the delayed posttest (36%, $\beta = 0.03$, $SE = 0.63$, $p = .966$), this group returns to an error rate very close to the one seen at the start of the study. The predicted error probability at each essay time for each group is presented in Table 23.

In comparison to the control group, only the increase in accuracy seen at Time 2 by the metalinguistic group was significantly different ($\beta = -1.632$, $SE = .76$, $p = .032$).

Table 23

Predicted Probabilities (log odds with percentage of errors) for Pronominal Duplication at Each Essay Time

	Pretest (Essay 1)	Essay 2 (after feedback on Essay 1)	Posttest (Essay 3)	Delayed Posttest (Essay 4)
Direct Feedback Group	-0.17 (46%)	-1.09 (25%)	-0.67 (34%)	-.84 (30%)
Metalinguistic Feedback Group	.27 (57%)	-1.11 (25%)	-0.30 (43%)	.00 (50%)
Control Group	-.59 (36%)	-.34 (42%)	-.16 (46%)	-.56 (36%)

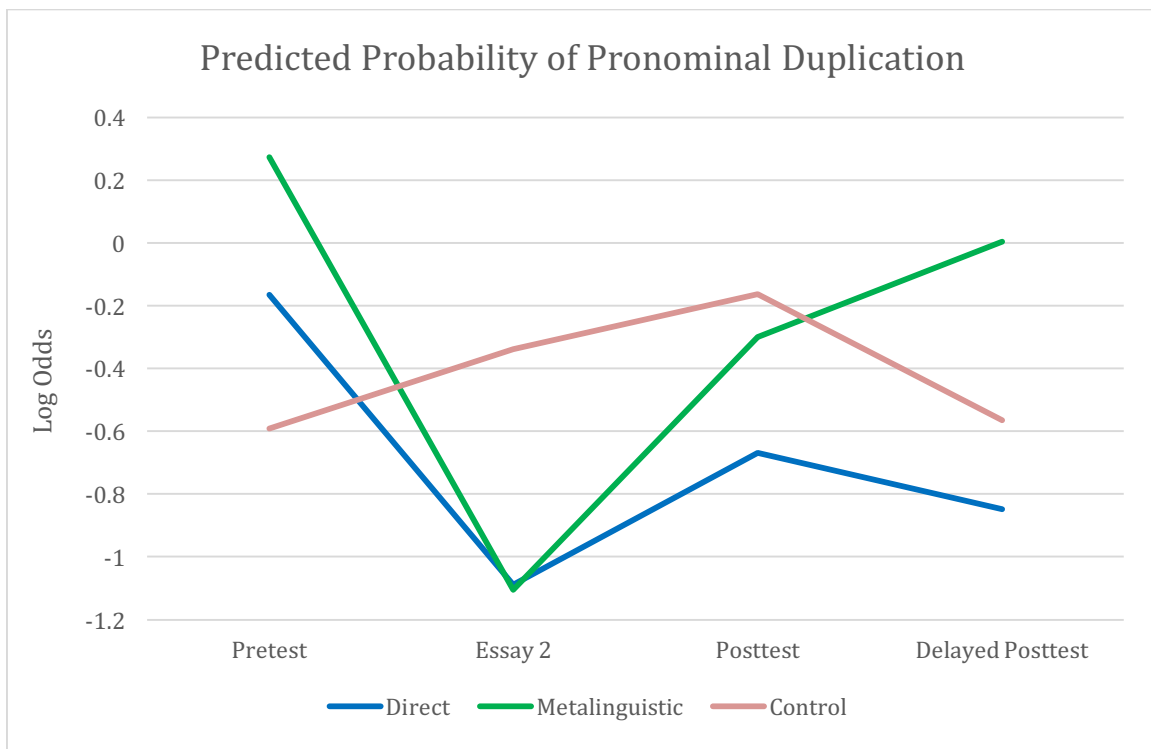


Figure 10. Plot of predicted probabilities for pronominal duplication by group at each essay time.

To summarize, both the direct and metalinguistic groups see improvement in accuracy with pronominal duplication from the time of the pretest to the time of the second essay; the improvement at Essay Time 2 is significant for the metalinguistic group, but not significant for the direct group. The gains in accuracy decrease for the two treatment groups by the time of the posttests. The control group showed no significant changes in accuracy at any point. As with the dropped pronoun analysis, these findings should be considered exploratory due to the small number of observations.

5.7 Summary of Findings from Mixed-effects Models

Table 24 presents a summary of significant gains in predicted probability of accuracy for each structure and for each group at the time of the written posttest and delayed posttest. These findings will be discussed in detail in the next chapter.

Table 24

Summary of Significant Gains in Accuracy on Immediate and Delayed Written Posttests

Structure	Direct Feedback		Metalinguistic Feedback		Control	
	Posttest	Delayed Posttest	Posttest	Delayed Posttest	Posttest	Delayed Posttest
Simple Past Tense	✓	✓	✓	✗	✗	✗
Present Perfect Tense	✓	✗	✓	✗	✗	✗
Dropped Pronouns	✗	✓	✓	✓	✗	✓
Pronominal Duplication	✗	✗	✗	✗	✗	✗

5.8 Qualitative Data from the Exit Survey

Participants completed the experimental protocol by answering a short background questionnaire (see Appendix C) to capture other factors that might provide additional insight into L2 performance. As shown in Figure 11, over half of the participants had no prior college experience before enrolling in the EAP program at St. Petersburg College. Although there are some older students in the program who may have completed several years of college prior to coming to the United States, it is likely that some of the students (e.g., one learner reported 14 years of college) did not understand this question.

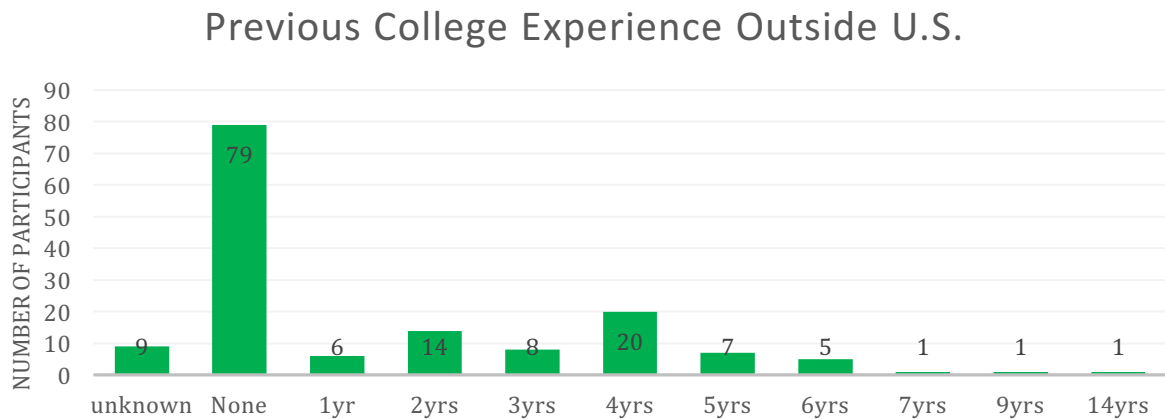


Figure 11. Participants' previous college experience outside U.S.

The participants were also queried regarding the feedback provided during the study and their feedback preferences in general. Figure 12 shows learner opinion by group concerning the feedback given in the current study, and we can see that even learners in the control group, who only received feedback on content, found the guidance to be positive.

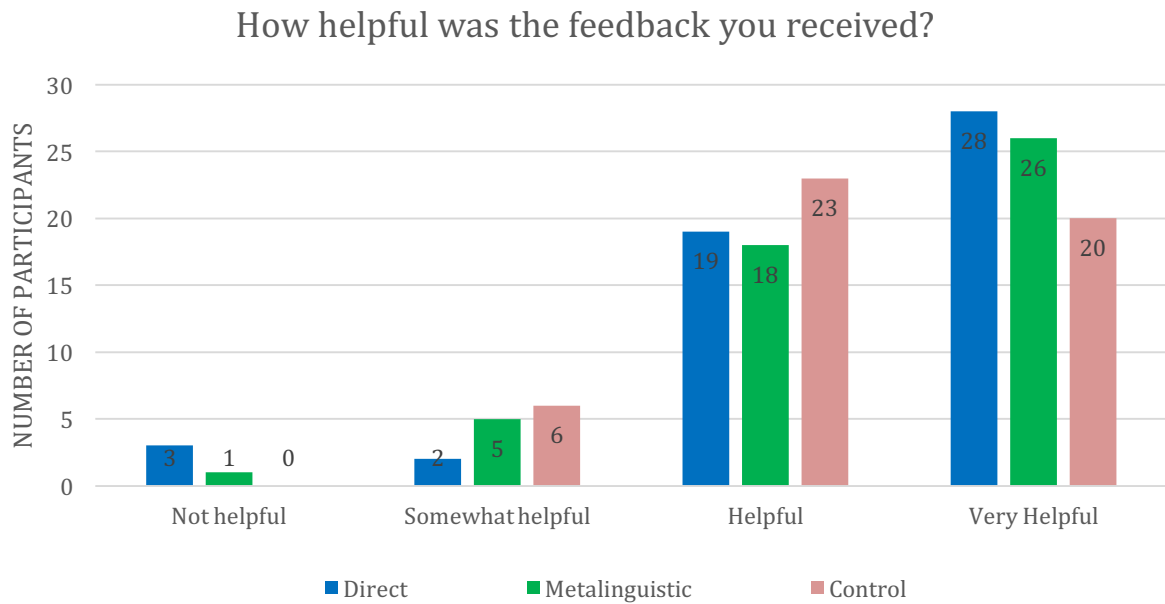


Figure 12. Learner opinion regarding feedback received.

Participants were also asked to indicate which types of feedback they preferred. The survey allowed them to select multiple responses. Research has shown that students believe strongly in the value of written CF for long-term improvement, and that students want comprehensive feedback on both ideas and grammar (e.g., Ferris, 1995; Ferris et al., 2013; Leki, 1991; Saito, 1994). Figure 13 clearly supports previous findings indicating learner preferences for feedback on both content and grammar. Although student preferences should not be the only factor considered by instructors when offering feedback, Bitchener and Ferris suggest that “consideration of student desires is a variable that should be in the mix” (2012, p. 95), and that due to great individual variation in learning, motivation, and proficiency, writing teachers should work together with their students to determine which types of feedback are best.

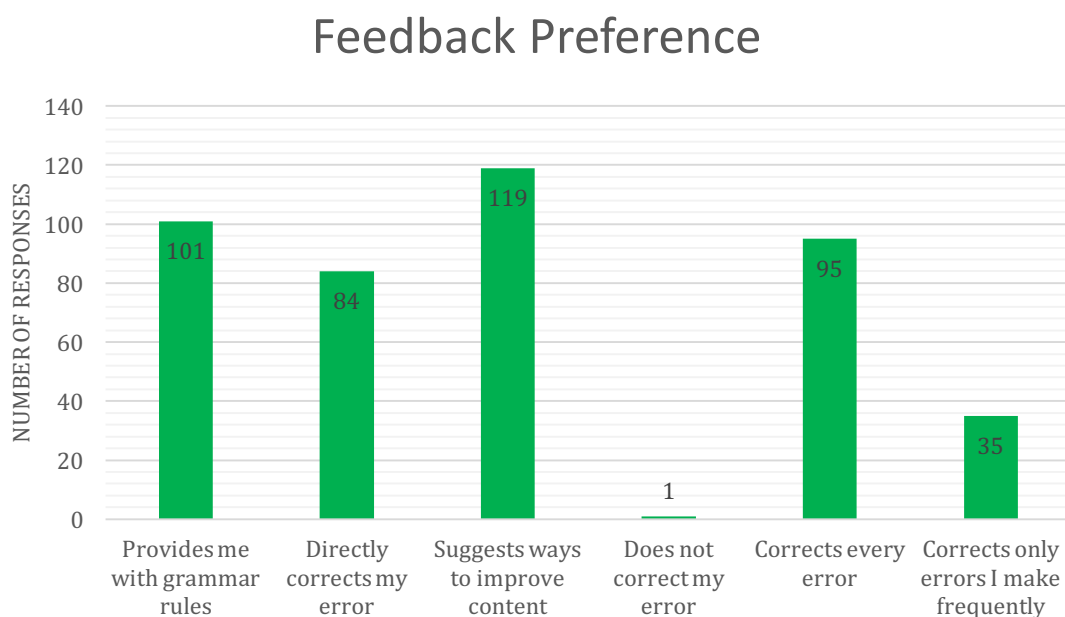


Figure 13: Learners’ feedback preferences.

Finally, learners were asked about their level of concern with grammatical accuracy. As previously discussed, motivation is an important factor that determines what learners do with the feedback they receive. Figure 14 shows that in this EAP setting, where learners were preparing to matriculate into regular college level courses, the majority of the learners were concerned about grammatical accuracy. However, forty-four were only “somewhat worried,” and seven individuals indicated they were “not worried at all.” Kormos (2012) argues that “both cognitive and motivational variables play an important role in goal-setting, determining the focus and amount of attention paid to the writing process and the feedback received, and the depth of problem-solving” (p. 393). Figures 13 and 14 combined, therefore, confirm that these students were motivated to write accurately and that they expected and desired corrective feedback. The scope of the current study did not allow for more detailed measures of

motivation, yet learners were asked to explain what they did with the feedback they received. The comments offered suggest a common theme in that learners felt that feedback helped them to pay attention and notice grammar errors so that they could avoid them in the future. Other students reported that the feedback helped them to understand the rules and that they studied these rules after class, using their grammar books, in an effort to avoid repeating the same mistakes. For example, one participant wrote, “after I got the corrections it make me more aware in my writing. I start read more carefully after I wrote.”

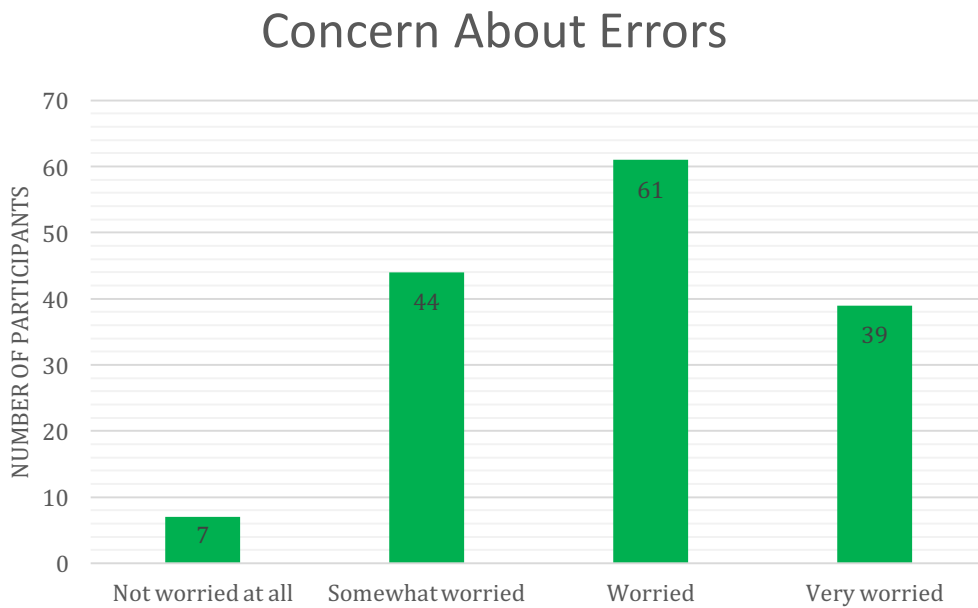


Figure 14. Learner concern about grammatical errors.

Chapter 6: Discussion and Conclusions

6.1 The Effects of Feedback on Accuracy in Subsequent Writing Tasks

The first two research questions concerned whether or not the provision of direct or metalinguistic feedback would lead to more accuracy on a new text than no feedback. In response to Truscott's (1996) claim that written corrective feedback is not helpful and that this practice should be abandoned, several carefully designed studies have now shown that written corrective feedback can improve accuracy in the writing of new texts over time (e.g., Bitchener, 2008; Bitchener & Knoch, 2008; Sheen, 2007; Shintani et al., 2014; Stefanou & Révész, 2015). Thus, the first two hypotheses, H1 and H2, predicted that when learners received two rounds of either focused direct feedback or focused metalinguistic feedback on the simple past tense, present perfect tense, dropped pronouns, and pronominal duplication, accuracy with these structures would improve on subsequent writing tasks, and that these learners would have greater gains in accuracy than learners in the control group, which did not receive feedback on linguistic errors.

As expected, at the time of the immediate posttest, both treatment groups showed progress with all four target structures compared to the control group, who did not show improvement with any structure. Increases in predicted accuracy were significant for both treatment groups for the simple past tense and the present perfect tense. The gains for pronoun dropping on the immediate posttest were only significant for the metalinguistic group; however, the direct group was trending in the correct direction with gains close to those predicted for the metalinguistic group. For redundant pronouns, the largest gains were seen on the second essay, after only one round of feedback. Specifically, the gains at essay time two for pronominal duplication were significant for the metalinguistic group

and marginal for the direct group in comparison to the control group, which showed no improvement with pronominal duplication. The gains enjoyed for this structure by the treatment groups, however, were not durable and no longer statistically significant at the time the immediate posttest, albeit these learners were still more accurate than at the start of the study. Given that only 21 learners made errors with pronominal duplication, these results should be considered exploratory and interpreted with caution. They do, however, suggest that both forms of feedback were enabling learners to recognize and lessen the number of redundant pronouns, whereas learners who did not receive feedback made more errors at times two and three and had shown no improvement by time four.

As far as long-term gains, the results are mixed. For the verb tenses, only learners who received direct feedback on the simple past maintained significant gains compared to the control group at the time of the delayed posttest. In contrast, learners who received metalinguistic feedback did not retain predicted improvements on the simple past tense by the time of the delayed posttest.

For the present perfect tense, neither treatment group maintained the gains shown on the immediate posttest when asked to write again four weeks later. That gains with the present perfect were not as durable as those made for the simple past tense is not surprising. Differences in treatability for each structure are discussed in the next section, 6.2.

Concerning long-term gains for pronoun errors, both the direct and metalinguistic groups exhibited a significant reduction in errors with dropped pronouns by the time of the delayed posttest; however, students in the control group, who demonstrated no improvement on the immediate posttest, had also significantly improved by this time. In contrast, the gains at the time of the delayed posttest made by both treatment groups were

no longer significant for errors with pronominal duplication, yet some improvements in predicted probability were retained. In other words, learners in the two treatment groups were still performing somewhat better with respect to pronominal duplication than at the start of the study, whereas the control group showed no change in accuracy from the first essay (36%) to the fourth essay (36%).

The fact that the treatment was not as durable long-term, as in some studies, may be because learners received feedback on two, three, or four structures and not just on one, as in many of the focused feedback studies. Comprehensive written feedback (unfocused) is common practice for many classroom teachers, yet as previously discussed, the results from studies which have examined the benefits of unfocused feedback are inconclusive. Van Beuningen et al. (2012) found long-term gains for unfocused feedback, but Frear (2012), for example, only saw long-term gains for focused feedback. Finding the magic number of errors to correct for each learner may never be determined, yet targeting only one error at a time is impractical in the classroom, where time is limited. Thus, providing feedback on four target errors gives this study more ecological validity. It may be the case, though, that long-term gains were not attainable after only two rounds of feedback since learners had to give their attention to several errors simultaneously, depending on their L1 and acquisition stage. Truscott (1996, 2001) also questions the likelihood that instructors can provide learners with feedback at just the right time. He concludes that “errors that are related to developmental sequences are not good targets for correction” (Truscott, 2001, p. 103). Ortega (2009) claims that “teachers cannot teach everything they want and instead should carefully consider what their students are developmentally ready to learn” (p. 99). Yet, she also states that learners who are instructed can progress at a faster rate and can progress further along

developmental sequences than naturalistic learners. For example, without instruction many naturalistic learners never produce “ed” or produce it with low levels of accuracy in comparison to those who receive instruction (e.g., Bardovi-Harlig, 1995). Ortega (2009) does not advise, however, trying to teach to sequences and processes since there are insufficient descriptions of the grammar for any target language. So what should be done? Bitchener and Knoch (2010a) conclude that determining which errors should be treated is not an impossible task for individual or small group contexts, such as those often found within second-language writing courses offered at many colleges or universities, since learners and teachers can work together to identify and target reoccurring and problematic forms. Additional longitudinal studies like the one conducted by Bitchener and Knoch (2010a), which examine slightly less focused feedback, are needed to determine whether or not additional rounds of feedback would lead to long-term gains when multiple structures are targeted.

In summary, the provision of written corrective feedback proved to be better than no feedback, especially in the short term. At the time of the immediate written posttest, both treatment groups displayed improved accuracy with all four errors in comparison to the control group. Some of the effects were more durable than others, and the control group was able to catch up to the other groups on pronoun dropping by time four, but the results for the pronoun dropping analysis are based on a very small group of participants and should be considered exploratory. Further research on the effectiveness of feedback for this structure is needed to confirm or reject these results.

On the whole, the present study, in contrast to Truscott’s claims (1996, 1999, 2001, 2004, 2007), corroborates the findings of other written corrective feedback studies (e.g., Bitchener, 2008; Bitchener & Knoch, 2008; Sheen 2007, 2010; Stefanou & Révész,

2015; Van Beuningen et al., 2012) that feedback on grammatical error can lead to learning and improved accuracy on subsequent writing tasks.

6.2 Error Type and Treatability

The third and fourth sets of hypotheses made predictions about the effectiveness of feedback for different structures. Both errors with verb tense and errors with pronouns were expected to be treatable by both direct and metalinguistic feedback. Nevertheless, some structures were expected to be more responsive to treatment than others.

First, regular verb forms were expected to be more treatable than irregular verb forms. Specifically, hypothesis H3a predicted that the provision of direct written CF would lead to greater gains in accuracy with the regular simple past and perfect forms than with irregular verb forms. For the same reasons, hypothesis H4a predicted that the provision of metalinguistic feedback would lead to greater gains with regular forms than with irregular forms. These predictions were based on the idea that regular verbs follow a pattern and are thus more treatable than irregular verbs for which there is no clear pattern. Frear (2012), for instance, found that the provision of written CF improved accuracy for regular verbs but not for irregular verbs. The results from the current study, however, do not support these hypotheses. There was no interaction effect found between verb type and essay time for any of the groups suggesting that the gains made with regular verbs were not significantly different from the gains made with irregular verbs. One explanation for this may be that due to the real-world nature of the writing tasks, many of the irregular verbs employed by the learners were high-frequency easily retrieved items, such as the past tense forms of the verbs “to be”, “to do”, “to have”, and “to get.”

However, many of the more advanced learners did use, and showed improvement with, low-frequency irregular verbs.

Second, the simple past tense was expected to be more treatable than the present perfect tense. Hypothesis H3b predicted that the provision of direct written CF would also lead to greater gains in accuracy with the simple past tense compared to the present perfect tense, which was considered more abstract. Similarly, hypothesis H4b predicted that the provision of metalinguistic feedback would result in greater gains in accuracy with the use of the simple past than with the use of the present perfect. Hypothesis H3b was confirmed since the gains made by learners who received direct feedback for the simple past tense were durable, whereas the gains made for the present perfect tense were not. Hypothesis H4b was not fully supported since learners who were given metalinguistic feedback achieved only short-term gains for both the simple past tense and the present perfect tense, but long-term gains for neither. An exact comparison of predicted gains for each structure would not be valid since the number of trials for the simple past tense was much greater overall than the number of trials for the present perfect. The results, however, indicate that the simple past error rate was almost cut in half from the time of the posttest, from 31% predicted errors to 16% for the direct group and from 20% to 12% for the metalinguistic group. For the present perfect tense, the error rate prediction was reduced from 70% at the time of the pretest for the direct group to 49% at the time of the posttest, and from 65% to 45% for the metalinguistic group. So, the predicted probability of error was not reduced as much as it was for the simple past, offering some support for the claim that feedback overall would be slightly more effective for the simple past tense than for the present perfect tense, since the perfect tense was predicted to be more abstract than the simple past. At the start of the study,

learners in all three groups were predicted to correctly use the simple past tense between 69% and 80% of the time, whereas they were only using the present perfect tense correctly between 30% and 42% of the time. Learners who had already experimented with present perfect before the provision of feedback would be expected to show greater increases in accurate use than learners who had not yet attempted to produce this form (Bardovi-Harlig, 1997). So, it is likely, that for learners with lower proficiency, additional practice and feedback with the present perfect tense is needed before durable gains can be maintained.

Finally, both types of pronoun errors were expected to be equally treatable. Hypothesis H3c predicted that the provision of direct written CF would be as effective for treating pronoun dropping as for redundant pronouns. And hypothesis H4c predicted that the provision of metalinguistic written CF would be equally effective in treating pronoun dropping and redundant pronouns. Although the results need to be interpreted with caution, the outcome scores suggest that learners who received direct feedback as well as learners who received metalinguistic feedback were able to reduce errors with both dropped pronouns and redundant pronouns. Whether or not the treatment was equally successful for both errors cannot be determined from this study due to the substantial difference in observations for each structure. Learners who received direct feedback were able to reduce their predicted error rate with dropped pronouns from 5% at the start to 1% at the time of the delayed posttest, and this reduction was significant. Similarly, learners who received metalinguistic feedback reduced their error rate from 6% at the start of the study to 3% at the end, and this change was also significant. By the time of the immediate posttest, accuracy had improved for both treatment groups, but not for the control group. The control group was able to catch up by the time of the delayed posttests. One

explanation for the improved performance of the control group on the delayed posttest is that some of the students in this group may have sought input on this structure during the weeks between the immediate and delayed posttest. Students in the EAP program are regularly encouraged to visit the writing center on campus for tutoring during the semester. As is always the case with classroom-based research, it is impossible to control for what happens outside of the classroom. Students were likely more concerned about all grammatical errors as the semester was drawing to a close because their grades would be finalized soon, and learners in this program are required to take a high-stakes final timed-writing exam. It is also possible that students in the treatment groups passed on information about the feedback they were receiving to the students in the control group.

As for errors with pronominal duplication, between time one and time two learners in the direct group made marginal gains in accuracy, and the metalinguistic group achieved significant gains. However, at time three, the error rates for both groups increased, and the gains in accuracy seen at time two were no longer significant at the time of the immediate posttest. Thus, learners in both treatment groups showed some improved accuracy with both pronoun errors; however, whether or not one error responded to treatment more than the other cannot be measured given the small and unequal number of participants in these analyses. Thus, it would not be prudent to confirm or reject H3c or H4c at this time. The results are inconclusive and should be considered exploratory. Further research is needed with a greater number of participants before any substantial claims can be made concerning the effectiveness of different feedback types on these two linguistic errors.

6.3 Interactions Between Feedback Type and Structure

The fifth set of hypotheses made predictions about the effectiveness of each type of feedback for each target error in order to determine if there was a differential effect of feedback type for different structures. In an effort to replicate and confirm findings, many recent written feedback studies have focused on a particular aspect of the English article system. Very few other structures have been investigated (exceptions include Shintani et al., 2014 and Diab, 2015). Also, to date, most of the focused written corrective feedback studies investigating different forms of direct feedback have compared direct feedback to direct supplemented with metalinguistic comments or instruction. Very few studies have considered direct feedback and metalinguistic feedback as individual predictors to see what merits each form of feedback has on its own for errors with other structures. One of the goals of the current study, therefore, was to examine these two forms of feedback separately. Of recent controlled studies, only Bitchener and Knoch (2010b), Diab (2015), Shintani and Ellis (2013), and Shintani et al. (2014) have considered the benefits of metalinguistic feedback or instruction on its own without the provision of the correct form. The results of these studies have suggested that metalinguistic feedback by itself can be as beneficial as direct feedback at least in the short term. Shintani and Ellis (2013) found short-term benefits for metalinguistic information on a new piece of writing shortly after the provision of feedback on English articles. Similarly, Shintani et al. (2014) found no difference between direct and metalinguistic feedback on a new writing task that occurred immediately following feedback for treating the hypothetical conditional. However, only Bitchener and Knoch (2010b) found long-term benefits for metalinguistic feedback. In contrast, several studies have found long-term benefits for direct feedback

(e.g., Bitchener, 2008; Bitchener & Knoch, 2008; Bitchener & Knoch, 2010a; Shintani et al., 2014).

Therefore, both types of feedback were expected to result in gains in accuracy for all of the target structures, at least in the short-term, but overall, direct feedback was expected to produce more durable gains for all four structures. Specifically, H5a stated that the provision of direct written CF would be more beneficial than the provision of metalinguistic feedback for treating errors with both the simple past tense and the present perfect tense regardless of whether the forms were regular or irregular. This hypothesis was only supported for the simple past tense.

When learners were given direct feedback, they significantly reduced their errors with the simple past tense. At the time of the pretest, this group was predicted to make errors with the simple past tense 31% of the time. However, by the time of the posttest, this group had reduced its error rate to 16%, almost 50% less. This gain in accuracy was maintained four weeks later on the delayed posttest. These results are in line with several studies that have examined the benefits of direct feedback on its own (without the addition of metalinguistic information) and have found promising long-term results. For example, Bitchener 2008 and Bitchener and Knoch (2008, 2010a) found that direct feedback on its own was just as good in the long term as direct feedback supplemented with metalinguistic explanation. The 2010a study was longitudinal and included three delayed posttests confirming that direct feedback was indeed durable for treating errors with English articles. Shintani et al. (2014) found that direct feedback that included an opportunity for revision was best long-term for treating the hypothetical conditional, and Stefanou et al. (2015) comparably found that direct on its own (even without revision)

was just as good as direct with metalinguistic for treating specific errors with English articles.

In the present study, metalinguistic feedback was not as durable as direct feedback for the simple past tense. Learners who received metalinguistic comments on their errors were predicted to make errors with the simple past tense 20% of the time at the start of the study. By the time of the immediate written posttest, this group had significantly reduced its error rate (almost in half like the direct group) and was only making errors with simple past tense 12% of the time. Yet, by the time of the delayed posttest, the metalinguistic group was producing errors 19% of time, essentially the same rate of error predicted at the start of the semester. This result agrees with the findings of Diab (2015) for pronoun agreement errors, Shintani and Ellis (2013) for errors with indefinite articles, and Shintani et al. (2014) for errors with the hypothetical conditional, all of whom found that metalinguistic feedback on its own resulted in significant gains in accuracy at the time of the immediate posttest, but that these improvements were lost by the time of the delayed posttest. In the present study, for the simple past tense, direct feedback was more durable than metalinguistic feedback and therefore considered more effective. Shintani et al. (2014) propose that direct corrective feedback offers learners positive evidence that allows them to confirm or reject their own attempts to use a structure, whereas metalinguistic information requires learners to apply an abstract set of rules to whatever errors they have produced. Learners in the present study likely had at least some knowledge of the rules for the simple past tense which potentially made the direct feedback more beneficial than the metalinguistic feedback for this structure.

One possible factor involved in the advantage of direct feedback for the simple past tense in the current study may be that learners were receiving feedback on both the

simple past and present perfect, which are related tenses. As is commonly found during language development, some learners began to overgeneralize the use of present perfect tense in cases where simple past was required; needless to say, this decreased their rate of correct use of the simple past. The metalinguistic feedback, which gave learners the rules to ponder, might have led to more overgeneralization than the direct feedback, which simply provided the correct form to copy.

For the present perfect tense, direct feedback was not superior to metalinguistic feedback. Both the direct feedback group and the metalinguistic group had significantly improved by the third writing task, which occurred after two rounds of feedback and revision. As expected, for the present perfect tense, the error rate was much higher at the start of the study for all learners compared to the error rate seen with the simple past. Learners in the direct group made errors with the perfect tense 70% of the time, meaning 30% of the time they used this verb tense correctly. By the time of the immediate posttest, this group was able to reduce its predicted error rate to 49%, meaning they were accurately using the perfect tense 51% of the time. By the time of the delayed posttest, the direct feedback group was only accurately using the present perfect tense 28% of the time (see Table 15). Similarly, learners who received metalinguistic feedback were predicted to make errors with the present perfect tense 65% of the time at the start of the study, indicating they were able to correctly produce this verb tense 35% of the time. This treatment group significantly reduced its error rate to 55% at the time of the immediate posttest, providing the present perfect tense 45% of the time. Like those who received direct feedback, however, this group was predicted to make errors 64% of the time four weeks later. So, the improvements in predicted accuracy were not durable for either treatment group. That direct feedback was not superior and effective for long-term

gains for the perfect tense might be due to developmental stage. Many learners may not have been ready to acquire this structure, whereas learners at all levels had likely already encountered the rules for the simple past and therefore had declarative knowledge that benefitted from production practice and explicit direct feedback that allowed them to confirm or reject their hypotheses concerning this structure.

Learners in the more advanced classes had likely received some explicit instruction concerning the rules for the present perfect tense, and therefore, had access to some declarative knowledge, while learners in the lower-level courses may not have encountered this tense before. It may be then that for the perfect tense, many learners needed both the rule explanation offered by metalinguistic feedback as well as the practice and confirmation offered by direct feedback, especially since they were receiving feedback on the simple past at the same time. Neither form of feedback was able to produce long-term gains with the present perfect on its own.

In summary, for the present perfect tense, both types of feedback enabled learners to significantly improve their accuracy on a new piece of writing, but neither group maintained this improvement long-term. Hypothesis H5a, therefore, was only supported for the simple past tense.

Finally, Hypothesis H5b predicted that the provision of direct written CF would have a greater effect on learners' accurate use of pronouns, both inclusion and exclusion, than the provision of metalinguistic information. This hypothesis was not supported. As previously noted, the number of learners who consistently made pronoun errors was very small ($n = 21$), so the results should be interpreted with care. We can see that for dropped pronouns, metalinguistic feedback appears to be somewhat more effective than direct feedback. At essay time two, after one round of feedback, the metalinguistic group is

already marginally better than at the start of the study. By time three, the immediate posttest, the metalinguistic group has significantly improved. This group was predicted to accurately supply subject pronouns 94% of the time at the start but was predicted to provide required pronouns 97% of the time by the time of the immediate posttest. This gain in accuracy was durable as learners continued to produce pronouns 97% of the time on a new writing task four weeks later. In contrast, the direct feedback group did not improve until the immediate posttest, and this improvement was not significant. Another finding that makes the pronoun dropping analysis problematic, however, is that by the time of the delayed posttest, all three groups had significantly improved, suggesting as previously discussed that there were other influential factors.

For pronominal duplication, both treatment groups trend in the direction of improved accuracy, yet the metalinguistic group sees a greater change in predicted accuracy than the direct feedback group at times two and three. At time two, the direct feedback group is only marginally improved, whereas the metalinguistic group is significantly better. Both groups lose some of their gains at the time of the immediate posttest (Time 3), but the metalinguistic group is still marginally better than at the start suggesting again that metalinguistic feedback was superior. Furthermore, only the gains achieved by the metalinguistic group are statistically different from those in the control group. In the long run, both feedback groups remain better than they were at the start, but the differences are no longer significant.

It may be that since pronouns are less salient and not as important in some cases for meaning as verb tense, metalinguistic feedback is necessary and more helpful than direct feedback that is simply copied during the revision process. In the case of the verb tenses, most learners had likely received form-focused instruction on at least the simple

past before the study began, and thus had some declarative knowledge of the rules that could be confirmed via direct feedback. In contrast, pronominal duplication and dropped pronouns only occurred regularly for learners whose L1s allowed this construction. Therefore, it is unlikely in an ESL classroom that instructors had given explicit instruction on these two errors at any point prior to the study. These learners had perhaps never been explicitly given the rule, and therefore lacked declarative knowledge about this error making the metalinguistic feedback more beneficial. Learners who received direct feedback only had to mechanically insert or delete a pronoun as indicated, and may not have been able to induce the rule.

6.4 Interactions Between Feedback Type and Language-Analytic Ability

The last set of hypotheses made predictions about the differential effect of language aptitude with different forms of feedback. Hypothesis H6a predicted that learners with a higher LAA would have greater overall gains in accuracy. This hypothesis was not fully supported. Learners with a higher LAA (one standard deviation above the mean) were predicted to outperform learners with an average or mean LAA in the direct feedback group, but this was only confirmed for one structure, the simple past tense. There was no correlation found between LLAMA F scores and accuracy scores for the present perfect tense; there was insufficient data to test for correlations between accuracy with pronouns and LLAMA F scores.

Surprisingly, learners with a higher LAA in the metalinguistic group did not have greater overall gains in accuracy with the simple past tense. Contrary to Hypothesis H6b, which predicted that learners with higher LAA would benefit more from the explanation than learners with lower LAA, the metalinguistic condition proved detrimental to learners

who had higher LLAMA F scores. At the time of the immediate posttest, learners with LLAMA F scores one standard deviation above the mean were predicted to receive essentially no benefit from the metalinguistic feedback compared to learners with an average LLAMA F score who significantly benefitted from rule provision. That there was no link found between individual differences in LAA and the gains made by the group who received metalinguistic comments agrees with the findings of Stefanou and Révész (2015), who only found a link between grammatical sensitivity and direct feedback, but contradicts the findings of Sheen (2007) who found a positive correlation between the provision of metalinguistic feedback and LAA.

Hypothesis H6c predicted that learners with a lower LAA would benefit more from the direct corrective feedback than they would from the metalinguistic explanation. This hypothesis was likewise not supported since learners with an average LLAMA F score, as well as learners with a LLAMA F score two standard deviations below the mean, received greater benefit in the metalinguistic group than learners with a higher LLAMA F score. Furthermore, both types of feedback led to significant improvements (for those with a mean LLAMA score) by the time of the immediate posttest. The hypothesis was based on the results of Sheen (2007)⁷ who found that learners with lower LAA were less likely to benefit from metalinguistic feedback. As previously discussed, Sheen (2007) explored how language-analytic ability correlated with uptake from direct correction concerning two aspects of English articles with or without metalinguistic feedback. In contrast to the current findings, learners in Sheen's study with higher LAA benefitted more from both feedback conditions compared to learners with lower levels of

⁷ The studies by Stefanou and Revesz (2015) and Shintani and Ellis (2015) were not published when the current study was designed and conducted.

LAA. However, unlike the treatment in the current study, the metalinguistic feedback was given in addition to direct feedback and not on its own, making an exact comparison impossible. Likewise, the structures examined in Sheen's study, English articles, were different from those examined in the current study.

As reviewed in Chapter 1, only two other written corrective feedback studies to date have investigated the interaction between language aptitude and written corrective feedback, and the results are mixed. Stefanou and Révész (2015) investigated the benefits of direct feedback compared to direct with metalinguistic on the accuracy of another aspect of English article use, generic and specific plural reference. As in the current study, and in contrast to Sheen (2007), participants with greater grammatical sensitivity and knowledge of metalanguage were more likely to achieve gains in the direct feedback group. There was no link found between individual differences and the gains made by the group who received direct feedback that included metalinguistic comments. It is highly likely that the state of these learners' metalinguistic knowledge played a role in their ability to make use of metalinguistic feedback. In other words, the role of LAA may also be mediated by other factors.

Shintani and Ellis (2015) found contrasting results with two different structures, the past hypothetical conditional and indefinite articles in their study, which compared direct feedback to metalinguistic explanation on its own. Like Sheen (2007), learners with greater LAA benefited more from both types of feedback than learners with weaker ability. However, in contrast to Sheen (2007), yet as was found in the current study, the mediating effect of LAA was only evident in new writing produced shortly after the feedback. Shintani and Ellis suggest that the extent to which LAA plays a role involves a complex interaction between type of feedback, opportunity to revise, and the target

structure. For accuracy with articles, LAA was more strongly correlated when learners in the direct feedback condition did not need to revise. The authors suggest that learners had to spend more time analyzing rather than just copying down the corrections. In the metalinguistic feedback condition, LAA was strongly implicated when learners were asked to revise since they had to figure out how to apply the rules. So for articles, LAA was implicated more when there was greater depth of processing demanded by the condition. In contrast, for the hypothetical conditional, the pattern was somewhat reversed. In the direct feedback condition, Shintani and Ellis speculate that revision led learners to draw more on their LAA, whereas when no revision was required, it did not. And, for the metalinguistic condition, there was a stronger correlation for LAA in the no revision condition perhaps because learners were able to focus on understanding the rule.

DeKeyser points out that

the interaction not only tells us about the importance of the internal (ID) variable, the external (treatment) variable, and their joint impact on the outcome variable, but also about the process that links them. This process is often treated as a black box, yet understanding it may be more important for generalizability of findings than endless (attempts at) empirical replication (failing to understand the interaction makes it hard to replicate findings; once the interaction is understood it becomes easier). (2012, p. 190)

Understanding the interactions between feedback type and individual differences, nevertheless, is tricky. Exact comparisons between the current studies that have examined the role of LAA in feedback effectiveness cannot be made. First, different measures of language aptitude were used in each study. Sheen's (2007) measure of aptitude was based on a language analysis test developed by Ottó and used by Schmitt, Dörnyei, Adolphs, and Durrow (2003). Shintani and Ellis (2015) used the language analytic ability section of the Language Aptitude Battery for Japanese (Sasaki, 1996), a Japanese translation of Pimsleur's Language Aptitude Battery. Stefanou and Révész (2015) employed a Words

in Sentences Test (adapted from the MLAT) and a test of metalanguage to measure grammatical sensitivity and knowledge of metalanguage respectively, and the current study used the LLAMA F. The aptitude measures used by Sheen (2007) and Shintani and Ellis (2015) required learners to induce the rules of an artificial language, but in Sheen's study learners had to determine the rules using English, their L2, whereas learners in the latter study could contemplate the rules using their L1. Moreover, the aptitude measure used by Stefanou and Révész required learners to understand the functions of words in sentences. Thus, the aptitudes being measured in these studies in relation to the treatments are not exactly the same, making the comparison of ATIs unclear.

Second, predictors were not operationalized in the same way across studies. Sheen (2007) and Stefanou and Révész (2015) both examined direct feedback versus direct feedback that included metalinguistic comments, whereas Shintani and Ellis (2015) and the current study compared direct feedback to metalinguistic on its own. Moreover, Shintani and Ellis (2015) did not indicate errors and only provided a metalinguistic handout in the L1 leaving learners to identify errors on their own, whereas the current study highlighted errors for students and provided metalinguistic rules in the margins using the L2. Finally, as previously discussed, the numbers of and types of structures varied across studies making the results difficult to compare. Although there can be no definitive answers at this time, there are some emerging trends that provide a glimmer of light inside the black box.

The most recent results from Shintani and Ellis (2015), Stefanou and Révész (2015), and the current study do not offer clear support to the view that LAA will play a stronger role when the feedback is metalinguistic in nature. The results do, however, lend some backing for the position that LAA is beneficial when learners have to work out the

grammar rules for themselves. The current finding that learners with a higher LAA received more benefit than learners with an average LAA in the direct feedback condition agrees with other studies that have found a strong correlation between more inductive instructional conditions and higher language aptitude (e.g., Erlam, 2005; Robinson, 1997).

Likewise, other inquiries have found that learners with lower LAA perform better with more explicit or deductive instruction. Hwu, Pan, and Sun (2014) investigated interactions between two types of explicit instruction and found that overall two explicit instructional approaches (deductive vs. explicit-inductive) had no differential effect on learning. However, learners with a low level of language aptitude performed significantly better with deductive instruction. Therefore, the metalinguistic feedback in the current study, which can be considered more explicit and deductive, should prove helpful for individuals with lower LAA.

That the higher aptitude learners did not improve with metalinguistic feedback is perplexing and more difficult to explain. Vatz et al. suggest that “interactions are often complex” (p. 285), so there could be additional interactions with proficiency level or with the target structures (e.g., in the current study, learners received feedback simultaneously on both the simple past and the present perfect) that have influenced the results. Both feedback conditions in the current study were explicit and direct. Certainly, the metalinguistic explanation with examples might be considered the more explicit in the sense that learners did not have to figure out the rules for themselves, and although the correct form was not explicitly given in this condition, examples were provided within the metalinguistic comments. The direct feedback was more inductive in that learners needed to understand which rule had been violated based on the correct form that was

provided. Other feedback studies (e.g., Shintani & Ellis, 2015) have likewise categorized direct feedback as inductive in the sense that learners must induce the rule from the correction provided. What must be considered then is whether or not the learner has any prior knowledge of the rule that could facilitate this process. In the current study, the simple past tense can be considered a relatively simple rule, and most of the learners in the study even at the lower course levels had already been introduced to this rule. Thus, in the case of the higher-aptitude learners, it may be that what was needed was practice that allowed these learners to accept or reject their formulations in order to test their own hypotheses about the rules making direct feedback more beneficial. In other words, higher-aptitude learners in the current study were likely to have some proceduralized knowledge of the rules for simple past tense and thus giving them the rules again was counterproductive, whereas the treatment which allowed them to confirm or reject their application or understanding of the rules was more beneficial. If indeed providing the rules was harmful for those with a higher LAA, this may in part explain why metalinguistic feedback on its own has been less successful overall in previous studies in comparison to direct feedback.

Future studies comparing these feedback types should include some measure of LAA to further test this interpretation. If direct feedback works as a leveler for some structures, and is good for both high and low aptitude types, then, metalinguistic on its own may not be the best choice. However, the combination of metalinguistic and direct feedback, provided within the text, as shown by Sheen (2007) and Diab (2015), may be beneficial and necessary for some learners and some structures.

6.5 Conclusions and Directions for Future Research

The current study sought to achieve two primary goals. First, this study aimed to contribute to the growing body of SLA research that has shown a positive relationship between written direct or metalinguistic error correction and improved, even retained accuracy on new pieces of writing over time. More importantly, this research intended to investigate interactions between LAA, feedback type, and four grammatical structures. Although several studies have now shown that focused corrective feedback is beneficial, few have considered grammatical errors other than English articles and interactions with LAA. The findings of the present study confirm the results of prior studies that have found written corrective feedback to be beneficial for different types of grammatical errors. Overall, learners in both feedback groups showed improvements in accuracy with all four structures in comparison to the control group, which showed no gains other than improvement with pronoun dropping on the delayed posttest. Figures 2, 5, 8, and 10 clearly illustrate that learners who received direct and metalinguistic feedback on their errors were able to improve their accuracy with the simple past tense, the present perfect tense, dropped pronouns, and redundant pronouns in comparison to those learners who did not receive feedback on their grammatical errors. Unlike studies with highly controlled writing tasks (e.g., Shintani et al., 2014; Stefanou & Revesz, 2015), one strength of the current study is that it included real-world writing tasks, like those required in the college's EAP and composition courses.

One limitation of the current study was the small number of participants who consistently dropped pronouns or included redundant pronouns. Variation in L1 is needless to say one of the drawbacks associated with research conducted in an ESL

environment. To examine the effect of feedback on structures specific to a given language, studies conducted in EFL settings are likely to provide more robust findings.

With respect to LAA, the present study aimed to explore whether or not language-analytic ability plays a role in determining which learners are more able to use written feedback to improve linguistic accuracy over time and for which structures. The results indicate a clear relationship between LAA and direct feedback, yet contrary to previous findings by Sheen (2007), metalinguistic feedback was not beneficial for learners with higher LAA. It may be that metalinguistic feedback that does not include direct correction has been less successful overall compared to direct feedback on its own because for some learners with higher LAA, requiring them to figure out the pattern or rule on their own and/or the confirmation offered by direct correction is more beneficial than rule provision.

Finally, this is the only written CF study, to the best of this author's knowledge, to examine the effect of providing learners of different proficiency levels with the same written corrective feedback, an area, as put forth by Bitchener and Knoch (2015), in need of research. Contrary to Truscott's claim (1996) that it is impossible to figure out where someone is at developmentally and therefore not practical to try to do so, this study has clear implications for instructed SLA. Direct feedback that requires revision may provide the production practice necessary to proceduralize declarative knowledge. The negative correlation found in the current study between the random intercept and the random slope for the simple past tense suggests that participants who began with less accuracy improved more than learners who started with greater accuracy. Furthermore, when pretest scores from the form-focused test were included in the model to account for variance in declarative knowledge, the model suggests that the provision of written

corrective feedback was effective for a mixed-level population. Teachers often have intuition about each student and know what errors are frequent, so it may be then that writing instructors can target a few errors at a time regardless of proficiency level. Although some have suggested that instruction (e.g., Ortega, 2009) and feedback (e.g., Truscott, 1996, 1999, 2001, 2004, 2007) may not alter the developmental route and could potentially be counterproductive or harmful by ignoring readiness, Ortega asserts that “it would be mistaken to conclude that instruction does not matter, just because it cannot override development” (p.100). Learners who receive instruction acquire faster and progress further and generally have greater accuracy compared to uninstructed learners (Long, 1983; Long, 1988; Ortega, 2009). Direct feedback, for instance, can provide input and allow learners to confirm or reject their own attempts with a structure, and metalinguistic feedback can provide rules for items that are less salient and often overlooked and thus harder to acquire without focused attention. As with oral forms of correction, written feedback on error allows learners an opportunity to notice the gap between their interlanguage and the target language, and to focus on form while revising a meaning-focused writing task.

As stated by Bitchener and Knoch (2015), “conclusions will continue to be elusive because each time different variables or combinations of variables are investigated, and the potential exists for such variables to mediate the findings in different ways” (p. 406). The current study offers important clues about the effectiveness of written CF for four linguistic targets. Individuals, even those in the same level ESL class, will vary in their starting level, in how they react to feedback, and in their ability to improve over time. Considering the large standard deviations often reported in written corrective feedback studies (e.g., Ferris, 2006), further research which includes a measure

of LAA and other measures of individual differences is needed to investigate the effects of different types of written feedback. Moreover, in order to better understand the complex relationship between LAA and written corrective feedback, future studies should explore interactions with other influential variables such as L1, salience, and proficiency level. Although many questions still remain, the findings of the present study confirm a positive role for written corrective feedback in instructed second language acquisition, at least for some learners and some structures.

Appendix A

Pretest Questions

Note: These tests were delivered electronically. Participants only saw one item at a time and were not able to backtrack. Time for the test was limited to 25 minutes. Items and answer choices within each section were delivered in random order.

Section 1: Choose the correct completion for each sentence.

1. Apple Computers _____ many different iPods since 2001.
a. produces b. produced c. has produced d. have produced
2. Juan _____ too much chocolate cake at the party last night.
a. eats b. ate c. has eaten d. have eaten
3. Leo _____ on the soccer team, the Switchbacks, for eight years. He will play with the Switchbacks again next year.
a. plays b. played c. has played d. have played
4. The Browns _____ in Korea for fifteen years. They live in Seoul, the capital.
a. live b. lived c. has lived d. have lived
5. Last week, I _____ my dog to the veterinarian.
a. take b. took c. has taken d. have taken
6. Tim _____ any good movies lately.
a. sees b. didn't see c. hasn't seen d. haven't seen
7. My parents _____ members of the swim club since 2002.
a. are b. were c. has been d. have been
8. Nathalie _____ me an email two weeks ago.
a. sends b. sent c. has sent d. have sent
9. David's daughter _____ from Stockholm University in 2008.
a. graduates b. graduated c. has graduated d. have graduated
10. The weather _____ especially warm and sunny lately.
a. is b. was c. has been d. have been

Section 2

1. Choose the grammatically correct sentence.
 - a. The girl who catches the flowers she is considered lucky.
 - b. The girl who catches the flowers is considered lucky.
 - c. The girl catches the flowers she is considered lucky.
 - d. Who catches the flowers she is considered lucky.
2. Choose the grammatically correct sentence.
 - a. People they can save energy by turn off lights.
 - b. People they can save energy by turning off lights.

- c. People can save energy by turning off lights.
 - d. People can save energy by turn off lights.
3. Choose the grammatically correct sentence.
 - a. The people they were horrified.
 - b. The people who watched the plane crash they were horrified.
 - c. The people who watched the plane crash were horrified.
 - d. The people when watched the plane crash were horrified.
 4. Choose the grammatically correct sentence.
 - a. The professor who teaches French travels to Europe every summer.
 - b. The professor who teaches French she travels to Europe every summer.
 - c. The professor who teaches French she travel to Europe every summer.
 - d. The professors who teaches French she travels to Europe every summer.
 5. Choose the grammatically correct sentence.
 - a. For example, pollution is a big problem today.
 - b. For example, pollution it is a big problem today.
 - c. For example, pollution a big problem today.
 - d. For example, is a big problem today.
 6. Choose the grammatically correct sentence.
 - a. Dishonesty is other problem.
 - b. Dishonesty it is another problem.
 - c. Dishonesty is another problem.
 - d. Dishonesty it another problem.
 7. Choose the grammatically correct sentence.
 - a. Restaurants prepares the food early in the day.
 - b. Restaurants they prepares the food early in the day.
 - c. Restaurants they prepare the food early in the day.
 - d. Restaurants prepare the food early in the day.
 8. Choose the grammatically correct sentence.
 - a. The man works at the restaurant likes to travel.
 - b. The man who works at the restaurant he likes to travel.
 - c. The man who works at the restaurant likes to travel.
 - d. The man who works at the restaurant travel.
 9. Choose the grammatically correct sentence.
 - a. For instance, pollutions is a big problem today.
 - b. For instance, pollution is a big problem today.
 - c. For instance, pollution big problem today.
 - d. For instance, pollution it is a big problem today.
 10. Choose the grammatically correct sentence.
 - a. New students they can apply for the scholarship.
 - b. New students he can apply for the scholarship.
 - c. New students can for the scholarship.
 - d. New students can apply for the scholarship.
 11. Choose the grammatically correct sentence(s).
 - a. When children want to learn is easier to teach them.
 - b. When children want to learn it is easier to teach them.
 - c. When children want to learn easier to teach them.
 - d. When want to learn it is easier to teach them.

12. Choose the grammatically correct sentence(s).
- Is obvious that the customers are not happy.
 - It is obvious that the customers are not happy.
 - It obvious that the customers are not happy.
 - Is obvious that they are not happy.
13. Choose the grammatically correct sentence(s).
- Last year Jose played soccer; however, he is too busy last year.
 - Last year Jose played soccer; however, too busy this year.
 - Last year Jose played soccer; however, he is too busy this year.
 - Last year Jose played soccer; however, is too busy this year.
14. Choose the grammatically correct sentence(s).
- After Samantha left work, she went to the store to buy food.
 - After Samantha left work, went to the store to buy food.
 - After Samantha left work, goes to the store to buy food.
 - After Samantha left work, she has been going to the store to buy food.
15. Choose the grammatically correct sentence(s).
- To be informed, very important to watch the news.
 - To be informed, was very important to watch the news.
 - To be informed, is very important to watch the news.
 - To be informed, it is very important to watch the news.
16. Choose the grammatically correct sentence(s).
- Pollution is bad for the environment, can cause problems for animals.
 - Pollution is bad for the environment. It causes problems for animals.
 - Pollution is bad for the environment, causes problems for animals.
 - Pollution is bad for the environment. It problems for animals.
17. Choose the grammatically correct sentence(s).
- Children learn because they interact with other people.
 - Children learn because interact with other people.
 - Children learn because they with other people.
 - Because they interact with other people.
18. Choose the grammatically correct sentence(s).
- The people were popcorn while watched the movie.
 - The people were eating while watched the movie.
 - The people were eating popcorn while watched the movie.
 - The people were eating popcorn while they watched the movie.
19. Choose the grammatically correct sentence(s).
- The men were talking when they were working.
 - The men were talking when were working.
 - The men were talking. Were working.
 - The men were talking when they working.
20. Choose the grammatically correct sentence(s).
- Men are strong. Can do difficult and dangerous jobs.
 - Men are strong. They difficult and dangerous jobs.
 - Men are strong. They can do difficult and dangerous jobs.
 - Men are strong. Can do and dangerous jobs.

Posttest Questions

Section 1: Choose the correct completion for each sentence.

1. Yuichi _____ for his vocabulary test for two hours yesterday.
a. has studies b. studies c. studied d. has studies
2. David likes to listen to music. Over the past two years, he _____ many songs on his computer.
a. downloaded b. has downloaded c. downloads d. has download
3. My wife and I _____ dessert after dinner on Saturday.
a. have shared b. sharing c. have been sharing d. shared
4. Mai _____ two great novels so far in her career.
a. has written b. writes c. wrote d. is writing
5. Nathan _____ me an email two weeks ago.
a. has sent b. sends c. is sending d. sent
6. Mr. and Mrs. Roberts _____ in Clearwater for five years, but now they live in Washington D.C.
a. lived b. have been living c. have lived d. live
7. The Martin family _____ in Florida for ten years. They live in Tallahassee, the capital of Florida.
a. lived b. are living c. have lived d. had lived
8. My brother repairs cars. He _____ on three different cars so far today.
a. works b. has worked c. worked d. has work
9. Melanie _____ a member of the International Club since 2010.
a. is being b. is c. was d. has been
10. Judy _____ her homework until 11 p.m. last night.
a. hasn't finished b. didn't finished c. didn't finish d. not finish

Section 2

1. Choose the grammatically correct sentence.
a. The man who won the lottery he is lucky.
b. The man winning the lottery he is lucky.
c. The man who won the lottery he lucky.
d. The man who won the lottery is lucky.
2. Choose the grammatically correct sentence.
a. Women they work hard at the factory.
b. Women work hard at the factor.
c. Women they working hard at the factory.
d. Women working hard at the factory.
3. Choose the grammatically correct sentence.
a. The people they were angry about the crime.
b. The people who saw the crime they were angry.
c. The people who saw the crime were angry.
d. The people when saw the crime were angry.

4. Choose the grammatically correct sentence.
 - a. The manager who works the night shift lives on Clearwater Beach.
 - b. The manager who works the night shift he lives on Clearwater Beach.
 - c. The manager he works the night shift and lives on Clearwater Beach.
 - d. The manager who working the night shift he is lived on Clearwater Beach.
5. Choose the grammatically correct sentence.
 - a. Crime it is a problem in big cities.
 - b. Crime being a problem in big cities.
 - c. Crime it being a problem in big cities.
 - d. Crime is a problem in big cities.
6. Choose the grammatically correct sentence.
 - a. Restaurants they are open early on the weekend.
 - b. Restaurants open early on the weekend.
 - c. Restaurants they open early on the weekend.
 - d. Restaurants early on the weekends.
7. Choose the grammatically correct sentence.
 - a. The train which stops in Tampa it will arrive soon.
 - b. The train which stops in Tampa will arrive soon.
 - c. The train which is stops in Tampa it will arrive soon.
 - d. The train which is stops in Tampa will arrive soon.
8. Choose the grammatically correct sentence.
 - a. Students who doing all of their homework they get good grades.
 - b. Students who is doing all of their homework get good grades.
 - c. Students who do all of their homework get good grades.
 - d. Students who do all their homework they get good grades.
9. Choose the grammatically correct sentence.
 - a. The teacher she wants all of the students to do well on the test.
 - b. The teacher she is want all of the students to do well on the test.
 - c. The teacher want all of the students did well on the test.
 - d. The teacher wants all of the students to do well on the test.
10. Choose the grammatically correct sentence.
 - a. The professor provided homework which it helped the students to review.
 - b. The professor provided homework which helped the students to review.
 - c. The professor provided homework which helping the students to review.
 - d. The professor he provided homework which helped the students to review.
11. Choose the grammatically correct sentence(s).
 - a. The fans left the stadium. Were sad that their team lost.
 - b. The fans left the stadium. They sad that their team lost.
 - c. The fans left the stadium. Was sad that their team lost.
 - d. The fans left the stadium. They were sad that their team lost.
12. Choose the grammatically correct sentence(s).
 - a. Maria worked at the hospital for many years. However, is retired now.
 - b. Maria worked at the hospital for many years. However, she is being retired now.
 - c. Maria worked at the hospital for many years. However, she is retired now.
 - d. Maria worked at the hospital for many years. However, is being retired now.

13. Choose the grammatically correct sentence(s).
- When the weather is cooler, is nice to play outside.
 - When the weather is cooler, it is nice to play outside.
 - When the weather is cooler, it nice to playing outside.
 - When the weather is cooler, is nice to playing outside.
14. Choose the grammatically correct sentence(s).
- After Christina went to the store, cooked dinner for her family.
 - After Christina went to the store, cooks dinner for her family.
 - After Christina went to the store, she cooked dinner for her family.
 - After Christina went to the store, she is dinner for her family.
15. Choose the grammatically correct sentence(s).
- In conclusion, is important to save energy.
 - In conclusion, important to save energy.
 - In conclusion, it is important to save energy.
 - In conclusion, it being important to save energy.
16. Choose the grammatically correct sentence(s).
- The student was unhappy because failed the test.
 - The student was unhappy because he failed the test.
 - The student was unhappy because he is failed the test.
 - The student was unhappy because fail the test.
17. Choose the grammatically correct sentence(s).
- People with money have too much power, and they can cause problems.
 - People with money have too much power, and causes problems.
 - People with money have too much power. Can cause problems.
 - People with money have too much power, and they causes problems.
18. Choose the grammatically correct sentence(s).
- The students were studying while were eating lunch.
 - The students were studying while they is eating lunch.
 - The students were study while were eating lunch.
 - The students were studying while they were eating lunch.
19. Choose the grammatically correct sentence(s).
- The children were smiling when were listen to the funny story.
 - The children were smiling when they be listening to the funny story.
 - The children were smiling when they were listening to the funny story.
 - The children were smiling when were listening to the funny story.
20. Choose the grammatically correct sentence(s).
- If the child is sick, can stay home from school.
 - If the child is sick, he can stay home from school.
 - If the child is sick, he staying home from school.
 - If the child is sick, can staying home from school.

Appendix B

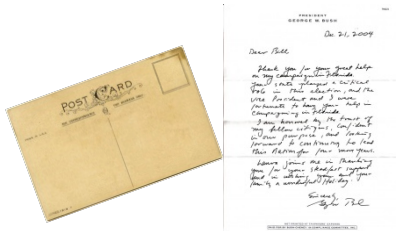
Writing Prompts

Write an essay in which you compare and contrast communication today to communication in the past. What has changed? Consider, for instance, changes, advancements or improvements in personal communication, business communication, and world news.

You will have one hour to write. Use your best grammar.

then

now



postcards
letters



electronic mail



rotary
phone



cell phone



black and white
television



flat screen television



print news



online news



face to face



social media



What has soccer star Lionel “Leo” Messi accomplished since he was born? Using the timeline and information below, write an essay (1-3 paragraphs) describing soccer player Lionel “Leo” Messi’s life and accomplishments up to now. Consider his career as a soccer player, his personal life, and his awards and other achievements.

You will have one hour to write. Use your best grammar.



1987 - born

1992 - begins playing football

2005 - first game for Barcelona FC (still on the team today)

2008 - gold medal with Argentina Olympic Team

2009 - serious relationship with partner/ girlfriend Antonella Rocuzzo

2010 - becomes goodwill ambassador for UNICEF (still an ambassador today)

2011 - permanent captain of the Argentina National Team

2012 - Messi’s son, Thiago, is born

2012 - highest paid on-field soccer player

2013 - meeting with Pope Francis



awards:

“World Soccer” Player of the Year - three times (2009, 2011, 2012)

ESPN “ESPY” Best International Athlete – two times (2012, 2015)

two World Cups (2009, 2011)

television advertisements:

Gillette

Turkish Airlines

Adidas

Pepsi

Write an essay which compares and contrasts education today to education 50 years ago. Discuss how education has changed over the past 50 years. Consider improvements in instructional techniques, classrooms, and materials.

You will have one hour to write. Use your best grammar.

then

now

library books



internet / databases



traditional books



electronic books



individual work



group work



1960's classroom



modern classroom

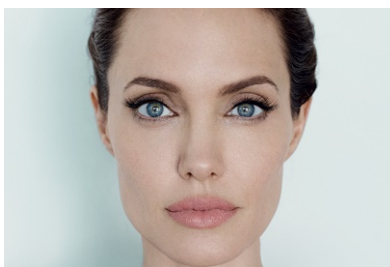


in-class instruction



on-line learning





What has actress Angelina Jolie accomplished since she was born? Using the timeline and information below, write an essay describing actress Angelina Jolie's life and accomplishments up to now. Consider her work, her awards, and her personal life (marriages and children).

You will have one hour to write. Use your best grammar.

1975 – born

1982 – first movie - *Lookin' to Get Out* (movie) as a child actor

1996 – marriage to Jonny Lee Miller (divorce in 1999)

1998 – Golden Globe Award – Best Supporting Actress in a Mini Series – *George Wallace*

2000 – Academy Award – Best Supporting Actress – *Girl, Interrupted*

2000 – marriage to Billy Bob Thornton (divorce in 2003)

2001 – Goodwill Ambassador for the United Nation's High Commissioner for Refugees

2002 – adoption of 7-month-old boy from Cambodia

2006 – birth of first baby, girl, in Namibia

2012 – special Envoy for the United Nations with a focus on refugee crisis

2014 – marriage to Brad Pitt (still married today)



Angelina and Brad (today) with their six children (3 adopted)

actress in 31 films including:

Mr. & Mrs. Smith

Wanted

Salt

Maleficent

The Tourist

humanitarian awards:

- Citizen of the World Award from the United Nation's Correspondents Association

- Global Humanitarian Award from the United Nation's Association

- Jean Hershoff Humanitarian Award from The Academy of Motion Pictures Arts and Sciences

- Honorary Dame, Commander of the order of St. Michael and St. George by Queen Elizabeth II



Appendix C

Exit Questionnaire: Participant Background Information and Debriefing Questions

1. Age: _____
2. Male _____ Female _____
3. Country of birth: _____
4. First (native) language: _____
5. Is this your first time attending college in the U.S.? Yes _____ No _____
6. Have you attended college in another country? If so, for how many years?
7. How long have you lived in the United States?
 - a. 0-6 months
 - b. 6 months – 1 year
 - c. 1-3 years
 - d. more than 3 years

8. How worried or concerned are you about errors in your writing?

0 = not worried at all
1 = somewhat worried
2 = worried
3 = very worried

9. How helpful was the feedback that you received on your errors?

10. I prefer feedback which (you may select more than one)
 - a. provides me with the grammar rules.
 - b. directly gives me the correct form or word.
 - c. suggests ways to improve the content of my writing.
 - d. does not correct my grammar.

11. Please explain briefly how you used the feedback you received to improve your writing.

References

- Ammar, A. & Spada, N. (2006). One size fits all? Recasts, prompts and L2 learning. *Studies in Second Language Acquisition*, 28, 543-574.
- Ashwell, T. (2000). Patterns of teacher response to student writing in a multi-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*, 9, 227-257.
- Ayoun, D. (2001). The role of negative and positive feedback in the second language acquisition of *passé composé* and *imparfait*. *Modern Language Journal*, 85, 226-243.
- Ayoun, D. (2004). The effectiveness of written recasts in the second language acquisition aspectual distinctions in French: A follow-up study. *The Modern Language Journal*, 88, 31-55.
- Bardovi-Harlig, K. (1995). The interaction of pedagogy and natural sequences in the acquisition of tense and aspect. In F. Eckman, D. Highland, P. Lee, J. Mileham, & R. Weber (eds.), *Second language acquisition theory and pedagogy* (pp. 151-168). Mahwah, NJ: Lawrence Erlbaum.
- Bardovi-Harlig, K. (1997). Another piece of the puzzle: The emergence of the present perfect. *Language Learning*, 47(3), 375-422.
- Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68 (3), 255-278.
- Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using

- S4 classes. R package version 0.999375-40. <http://CRAN.R-project.org/package=lme4>
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing, 17*, 102–118.
- Bitchener, J. & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. New York: Routledge.
- Bitchener, J. & Knoch, U. (2008). The value of written corrective feedback for migrant and international students. *Language Teaching Research, 12*, 409-431.
- Bitchener, J. & Knoch, U. (2010a). The contribution of written corrective feedback to language development: A ten month investigation. *Applied Linguistics, 31*, 193-214.
- Bitchener, J. & Knoch, U. (2010b). Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing, 19*, 207-217.
- Bitchener, J. & Knoch, U. (2015). Written corrective feedback studies: Approximate replication of Bitchener & Knoch (2010a) and Van Beuningen, De Jong & Kuiken (2012). *Language Teaching, 28* (3), 405-414.
- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing, 14*, 191-205.
- Celce-Murcia, M. & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course*. (2nd Ed.). Boston: Heinle & Heinle.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in

- the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12, 267-296.
- DeKeyser, R. (1993). The effect of error correction on L2 grammar knowledge and oral proficiency. *Modern Language Journal*, 77, 501-514.
- DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty & J. Williams (Eds.), *Focus on form in second language classroom acquisition* (pp. 42-63). New York: Cambridge University Press.
- DeKeyser, R. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, 55(S1), 1-25.
- DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning*, 62 (Suppl. 2) 189-200.
- DeKeyser, R. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition. An introduction* (pp. 94-112). London: Routledge.
- DeKeyser, R. (in press). Of moving targets and chameleons: Why the concept of difficulty is so hard to pin down. *Studies in Second Language Acquisition*.
- DeKeyser, R., & Koeth, J. (2011). Cognitive aptitudes for second language learning. In E. Hinkle (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 395-406). London: Routledge.
- Diab, N. (2015). Effectiveness of written corrective feedback: Does type of error and type of correction matter? *Assessing Writing*, 24, 16-34.
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in*

- second language acquisition* (pp. 31-64). Mahwah, NJ: Erlbaum.
- Dörnyei, Z. & Skehan, P. (2003). Individual Differences in L2 Learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Oxford: Blackwell.
- Doughty, C. (2013). Optimizing post-critical-period language learning. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 153-175). Amsterdam: Benjamins.
- Doughty, C. & Williams, J. (1998). Pedagogical choices in focus on form. In C. Doughty & J. Williams (Eds.). *Focus on form in classroom second language acquisition*. Cambridge: Cambridge University Press.
- Ellis, R. (2010). A framework for investigating oral and written corrective feedback. *Studies in Second Language Acquisition*, 32, 335-349.
- Ellis, R. (2009). A typology of written corrective feedback types. *English Language Teaching Journal*, 63, 97-107.
- Ellis, R., Sheen, Y., Murakami, M. & Takashima, H. (2008). The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System*, 36, 353-371.
- Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, 9(2), 147-171.
- Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178-190). Cambridge: Cambridge University Press.
- Ferris, D. R. (1995). Student reactions to teacher response in multiple-draft composition

- classrooms. *TESOL Quarterly*, 29, 33-53.
- Ferris, D. R. (1997). The influence of teacher commentary on student revisions. *TESOL Quarterly*, 31, 315-339.
- Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8, 1–10.
- Ferris, D. R. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Lawrence Erlbaum.
- Ferris, D. R. (2004). The “Grammar Correction” debate in L2 writing: Where are we, and where do we go from here? (and what do we do in the meantime...?). *Journal of Second Language Writing*, 13, 49-62.
- Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (eds.), *Feedback in second language writing: Contexts and issues*. Cambridge: Cambridge University Press, 81-104.
- Ferris, D.R. (2010). Second language writing research and written corrective feedback in SLA. *Studies in Second Language Acquisition*, 32, 181-201.
- Ferris, D.R., & Hedgcock, J.S. (1998/2004). *Teaching ESL composition: Purpose, process, and practice*. Mahwah, NJ: Lawrence Erlbaum.
- Ferris, D.R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, 22, 307-329.
- Ferris, D.R., & Roberts, B. (2001). Error feedback in L2 writing classes: how explicit does it need to be? *Journal of Second Language Writing*, 10, 161-184.
- Frear, D. (2012). *The effect of written feedback and revision on intermediate Chinese*

- learners' acquisition of English*. Unpublished doctoral dissertation. The University of Auckland, New Zealand.
- Gass, S., & Selinker, L. (2001). *Second language acquisition: An introductory course*. Mahwah, N.J.: L. Erlbaum Associates.
- Goldstein, L. (2006). Feedback and revision in second language writing: Helping learners become independent writers. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 185–205). New York: Cambridge University Press.
- Guénette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback in writing. *Journal of Second Language Writing*, 16(1), 40-53.
- Han, Z. (2002). A study of the impact of recasts on tense consistency in L2 output. *TESOL Quarterly*, 36 (4), 543-572.
- Havranek, G., & Cesnik, H. (2001). Factors affecting the success of corrective feedback. *EUROSLA Yearbook*, 1, 99-122.
- Hwu, F., Pan, W., & Sun, S. (2014). Aptitude-treatment interaction effects on explicit rule learning: a latent growth curve analysis. *Language Teaching Research*, 18(3), 294-319.
- Hyland, F. (2011). The language learning potential of form-focused feedback on writing: Students' and teachers' perceptions. In R. M. Manchón (Ed.), *Learning-to-write and writing-to-learn in an additional language* (pp. 159-180). Amsterdam: John Benjamins.
- Ishida, M. (2004). Effects of recasts on the acquisition of the aspectual form *-te i-(ru)* by learners of Japanese as a foreign language. *Language Learning*, 54(2), 311-394.

- Jaeger, F. (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–46.
- Kang, E., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal*, 99 (1), 1-18.
- Kormos, J. (2012). The role of individual differences in L2 writing. *Journal of Second Language Writing*, 21, 390-403.
- Kormos, J. (2013). New conceptualizations of language aptitude in second language attainment. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 131-152). Amsterdam: Benjamins.
- Lalande, J. F. II (1982). Reducing composition errors: An experiment. *Modern Language Journal*, 66, 140-149.
- Leki, I. (1990). Coaching from the margins: Issues in written response. In B. F. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 57-68). New York: Cambridge University Press.
- Leki, I. (1991). The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals*, 24, 203-218.
- Long, M. H. (1983). Does second language instruction make a difference? *TESOL Quarterly*, 17, 359-382.
- Long, M. H. (1988). Instructed interlanguage development. In L. M. Beebe (Ed.), *Issues in second language acquisition: Multiple perspectives* (pp. 115-141). New York: Harper and Row.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of second language*

- acquisition* (pp. 413-468). San Diego, CA: Academic Press.
- Long, M. H. (2009). Methodological principles for language teaching. In M. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 373-394). Malden, MA: Blackwell.
- Long, M. H. (2015). Towards a cognitive-interactionist theory of instructed SLA (ISLA). In M. H. Long (Ed.), *Second language acquisition and task-based language teaching*. Malaysia: Wiley-Blackwell.
- Mackey, A., & Oliver, R. (2002). Interactional feedback and children's L2 development. *System*, 30, 459–477.
- Mackey, A., & Philp, J. (1998). Conversational interaction in second language development: Recasts, responses, and red herrings? *The Modern Language Journal*, 82, 338–356.
- Meara, P. (2005). LLAMA language aptitude tests: The manual. *Swansea: Lognostics*.
- Ortega, L. (2009). Sequences and processes in language learning. In Long, M.H., & Doughty, C.J. (eds.), *Handbook of language teaching* (pp. 81-105). Oxford: Blackwell.
- Pienemann, M. (1998). *Language processing and second language development: Processability theory*. Amsterdam/Philadelphia: John Benjamins.
- Pienemann, M. (2005). *Cross-Linguistic aspects of Processability Theory*. Amsterdam: John Benjamins.
- Pienemann, M. (2015). An outline to processability theory and its relationship to other

- approaches to SLA. *Language Learning*, 65 (1), 123-151.
- Pienemann, M., & Keßler, J.-U. (2011). *Studying processability theory: Introductory textbook*. Amsterdam: John Benhamins.
- Pienemann, M., & Keßler, J.-U. (2012). Processability theory. In S. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 228-246). Abingdon, UK: Routledge.
- Polio, C., Fleck, C., & Leder, N. (1998). "If only I had more time": ESL learners' changes in linguistic accuracy on essay revisions. *Journal of Second Language Writing*, 7, 43-68.
- R Development Core Team (2011), R: A Language and Environment for Statistical Computing. Vienna, Austria : the R Foundation for Statistical Computing. ISBN: 3-900051-07-0. Available online at <http://www.R-project.org/>.
- Reid, J. (1994). Responding to ESL students' texts: The myths of appropriation. *TESOL Quarterly*, 28, 273-292.
- Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, 47, 45-99.
- Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research*, 17, 368-392.
- Sachs, R., & Polio, C. (2007). Learners' use of two types of written feedback on a L2 writing task. *Studies in Second Language Acquisition*, 29, 67-100.
- Saito, H. (1994). Teachers' practices and students' preferences for feedback on second language writing: A case study of adult ESL learners. *TESL Canada Journal*,

11(2), 46-70.

- Sasaki, M. (1996). *Second language proficiency, foreign language aptitude and intelligence*. New York: Lang.
- Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York: Cambridge University Press.
- Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker, *Proceedings of CLaSIC 2010*, Singapore, December 2-4 (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.
- Schmitt, N., Dörnyei, Z., Adolphs, S., & Durow, V. (2003). Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt (Ed.), *The acquisition, processing, and use of formulaic sequences* (pp. 55-86). Amsterdam: John Benjamins.
- Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, 41, 255–283.
- Sheen, Y. (2010). Differential effects of oral and written corrective feedback in the ESL classroom. *Studies in Second Language Acquisition*, 32, 203-234.
- Sheen, Y., Wright, D., & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*, 37(4), 556-569.

- Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, 22, 286-306.
- Shintani, N., & Ellis, R. (2015). Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing? *System*, 49, 110-119.
- Shintani, N., Ellis, R., & Suzuki, W. (2014). Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*, 64 (1), 103-131.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Stefanou, C., & Révész, A. (2015). Learner differences, and the acquisition of second language article use for generic and specific plural reference. *The Modern Language Journal*, 99 (2), 263-282.
- Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems*. Cambridge: Cambridge University Press.
- Trofimiovitch, P., Ammar, A., & Gatbonton, E. (2007). How effective are recasts? The role of attention, memory, and analytic ability. In A. Mackey (Ed.), *Conversational interaction in second language acquisition: A series of empirical studies* (pp. 171-195). Oxford: Oxford University Press.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, 327-369.

- Truscott, J. (1999). The case for “the case for grammar correction in L2 writing classes”: A response to Ferris. *Journal of Second Language Writing*, 8, 111–122.
- Truscott, J. (2001). Selecting errors for selective correction. *Concentric: Studies in English Literature and Linguistics*, 27 (2), 93-108.
- Truscott, J. (2004). Evidence and conjecture: A response to Chandler. *Journal of Second Language Writing*, 13, 337-343.
- Truscott, J. (2007). The effect of error correction on learners’ ability to write accurately. *Journal of Second Language Writing*, 16, 255-272.
- Truscott, J. & Hsu, A. Y-P. (2008). Error correction, revision, and learning. *Journal of Second Language Writing*, 17, 292-305.
- Van Beuningen, C., de Jong, N. H., & Kuiken, F. (2008). The effect of direct and indirect corrective feedback on L2 learners’ written accuracy. *ITL International Journal of Applied Linguistics*, 156, 279-296.
- Van Beuningen, C., de Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in Dutch multilingual classrooms. *Language Learning*, 62 (1), 1-41.
- Vatz, K., Tare, M., Jackson, S. R., & Doughty, C. J. (2013). Aptitude-treatment interaction studies in second language acquisition. In G. Granena & M. H. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 273-292). Amsterdam: Benjamins.
- Yang, Y. & Lyster, R. (2010). Effects of form-focused practice and feedback on Chinese EFL learners’ acquisition of regular and irregular past tense forms. *Studies Language in Second Language Acquisition*, 32, 235-263.

Yilmaz, Y. (2013). The role of working memory capacity and language analytic ability in the effectiveness of explicit correction and recasts. *Applied Linguistics*, 34 (3).